Master's Thesis

# CRISPRa mediated pluripotency reprogramming of biobanked lymphoblastoid cell lines

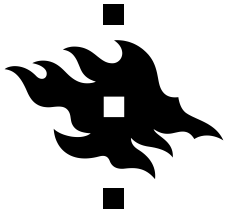
Laura Laiho

Molecular Biosciences

Faculty of Biological and Environmental Sciences

UNIVERSITY OF HELSINKI

June 2018



HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

| Faculty | Degree Programme |
|---|---|
| Faculty of Biological and Environmental Sciences | Molecular Biosciences |

Author
Laura Oona Annukka Laiho

Title
CRISPRa mediated pluripotency reprogramming of biobanked lymphoblastoid cell lines

Subject/Study track
Biotechnology

| Level | Month and year | Number of pages |
|---|---|---|
| Master's thesis | June 2018 | 67 |

Abstract

Lymphoblastoid cell line (LCL) collections offer a vast source of biological data of various diseases stored in biobanks. Even though LCLs have been used as a conducive surrogate in vitro cell model to study various diseases, it has been shown that they cannot recapitulate all the regulatory properties of a disorder specific primary tissue, constraining the utility of the collections. The rise of induced pluripotent stem cell (iPSC) technology has offered a novel strategy for repurposing the abundant LCL collections for regenerative medicine, drug discovery and disease modeling. However, the cell-type-specific reprogramming events in LCLs are not understood, hindering the development of a more efficient and reliable reprogramming method for these cell lines. To systematically study the endogenous events taking place during the reprogramming process, the CRISPR/Cas mediated gene activation (CRISPRa) platform has been harnessed in fibroblasts.

This thesis work aimed to establish the novel CRISPRa mediated reprogramming method for the first time in biobanked LCLs, obtained from the Psychiatric Family Collections of THL Biobank. The method relies on the activation of the endogenous pluripotency factors with CRISPR activators, which results in the induction of the reprogramming process. The screening of different factor combinations revealed that the successful reprogramming of LCLs seems to require the targeting of the EEA motif, a genomic element which is enriched in promoter regions of genes related to embryo genome activation. The reprogramming efficiencies were however varying, and a gRNA function validation assay showed that the gRNAs used in this project were suboptimal for efficient gene activation in LCLs. The reprogramming method was nevertheless demonstrated to successfully convert five biobanked LCLs into bona fide iPSC lines.

The iPSC lines generated in this thesis serve as a proof of concept that the novel CRISPRa reprogramming method works in biobanked LCLs and could be developed to systematically study the cell-type-specific reprogramming events in these cells. Also, the generated iPSC line collection constitutes the first iPSC collection derived from biobanked cells with the CRISPRa reprogramming method and is available as a future research resource, thus paving the way for the wider use of the LCL collections.

Keywords
CRISPR/Cas, CRISPRa, reprogramming, lymphoblastoid cell lines, induced pluripotent stem cells, iPSCs

Supervisor or supervisors
Professor Timo Otonkoski, docent Ras Trokovic

Where deposited
Viikki Science library, University of Helsinki

Additional information

| Tiedekunta | Koulutusohjelma | |
|---|---|---|
| Bio- ja ympäristötieteellinen tiedekunta | Molekyylibiotieteet | |
| Tekijä | | |
| Laura Oona Annukka Laiho | | |
| Työn nimi | | |
| Lymfoblastoidisolulinjojen CRISPRa välitteinen pluripotenssiuudelleenohjelmointi | | |
| Oppiaine/Opintosuunta | | |
| Biotekniikka | | |
| Työn laji | Aika | Sivumäärä |
| Pro gradu | Kesäkuu 2018 | 67 |

Tiivistelmä


Biopankkeihin tallennetut lymfoblastoidisolulinjakokoelmat tarjoavat laajan biologisen datalähteen monien sairauksien tutkimiseen. Näitä verisoluista peräisin olevia solulinjoja on käytetty menestyksekkäästi tautisolumallina useissa *in vitro* -tutkimuksessa. Ne eivät kuitenkaan pysty täysin toistamaan sairausspesifisten primäärisolutyyppien kaikkia säätelyominaisuuksia, mikä on rajoittanut kokoelmien käytön vain tietynlaiseen tutkimukseen. Indusoituihin monikykyisiin kantasoluihin (iPS-soluihin) perustuva teknologia on kuitenkin tarjonnut uudenlaisen strategian kokoelmien monipuolisempaan hyödyntämiseen uusiutuvan lääketieteen, lääkekehityksen ja sairausmallinnuksen saralla. Lymfoblastoidisolujen solutyyppispesifisiä uudelleenohjelmointitapahtumia ei kuitenkaan tunneta tarkasti, mikä hidastaa uusien tehokkaampien ja luotettavimpien uudelleenohjelmointimenetelmien kehittämistä. CRISPR/Cas välitteinen geeniaktivaatio (CRISPRa) saattaa tarjota tähän uusia työkaluja, sillä se on jo valjastettu endogeenisten uudelleenohjelmointitapahtumien systemaattiseen tutkimiseen fibroblasteissa.

Tämän Pro gradu -tutkielman tavoitteena oli osoittaa ensimmäistä kertaa CRISPRa välitteisen uudelleenohjelmointimenetelmän toimivuus THL Biopankin psykiatrisesta perhekokoelmasta saaduissa lymfoblastoidisolulinjoissa. Menetelmä perustuu solujen endogeenisten pluripotenssitekijöiden aktivointiin CRISPR-aktivaattoreilla, joka johtaa uudelleenohjemointiprosessiin soluissa. Eri pluripotenssitekijäkombinaatioiden joukosta löydettiin tehokkain yhdistelmä, ja tulosten perusteella näiden solulinjojen uudelleenohjelmointi näytti vaativan EEA-motiivin aktivoinnin. EEA-motiivi on genominen elementti, joka on rikastunut alkion genomiaktivaatioon liittyvien geenien promoottorialueille. Uudelleenohjelmointitehokkuus oli kuitenkin vaihteleva, ja osittaiseksi syyksi nähtiin projektissa käytettyjen opas-RNA:iden huono aktivaatioteho tässä solutyypissä. Menetelmä osoitettiin kuitenkin toimivaksi näissä soluissa, sillä viisi eri lymfoblastoidisolulinjaa saatiin onnistuneesti uudelleenohjelmoitua iPS-solulinjoiksi.

Tässä projektissa tuotetut iPS-solulinjat ovat osoitus siitä, että CRISPRa välitteinen uudelleenohjelmointimenetelmä toimii biopankkeihin tallennetuissa lymfoblastoidisolulinjoissa, ja tulevaisuudessa menetelmää voidaan kehittää solutyyppispesifien uudelleenohjelmointitapahtumien systemaattiseen tutkimiseen. iPS-solulinjakokoelma on lisäksi ensimmäinen CRISPRa menetelmällä tuotettu kokoelma, joka on saatavilla tutkimukseen THL biopankista, näin mahdollistaen lymfoblastoidisolulinjakokoelmien laajemman tutkimuskäytön.



Avainsanat
CRISPR/Cas, CRISPRa, uudelleenohjelmointi, lymfoblastoidisolulinjat, indusoidut monikykyiset kantasolut, iPS-solut

Ohjaaja tai ohjaajat
Professori Timo Otonkoski, dosentti Ras Trokovic

Säilytyspaikka
Viikin tiedekirjasto, Helsingin yliopisto

Muita tietoja

# Contents

# Abbreviations

| | |
|---|---|
| AP | alkaline phosphatase |
| Cas | CRISPR associated protein |
| cDNA | complementary DNA |
| CRISPR | clustered regularly interspaced short palindromic repeats |
| CRISPRa | CRISPR mediated gene activation |
| DSB | double stranded break |
| E8 | Essential 8 medium |
| EB | embryoid body |
| EBNA-1 | EBV nuclear antigen 1 |
| EBV | Epstein-Barr virus |
| EEA | EGA enriched Alu element motif |
| EGA | embryonic genome activation |
| ESC | embryonic stem cell |
| FACS | fluorescence activated cell sorting |
| GFP | green fluorescent protein |
| gRNA | guide-RNA |
| HFF | human foreskin fibroblast |
| H3K27ac | histone H3 lysine 27 acetylation marker |
| H3K27me3 | histone H3 lysine 27 trimethylation marker |
| ICC | immunocytochemistry |
| ID PCR | identification genotyping with polymerase chain reaction |
| iPSC | induced pluripotent stem cell |
| KLF4 | Krüppel-like factor 4 |

| | |
|---|---|
| LCL | lymphoblastoid cell line |
| MET | mesenchymal-to-epithelial transition |
| MG | matrigel |
| miRNA | micro-RNA |
| mRNA | messenger RNA |
| NaB | sodium butyrate |
| OCT4 | octamer binding protein 4 |
| OriP | origin of plasmid replication |
| OSKM | OCT4, SOX2, KLF4, C-MYC |
| PAM | protospacer adjacent motif |
| PCR | polymerase chain reaction |
| PDL | population doubling level |
| PP | pluripotency |
| PSC | pluripotent stem cell |
| RE | regulatory element |
| ROCK | Rho-associated protein kinase |
| RT | room temperature |
| RTase | reverse transcriptase |
| RT-qPCR | reverse transcription quantitative polymerase chain reaction |
| SAM | synergistic activation mediator |
| SEM | standard error of the mean |
| shp53 | silencing hairpin RNA for p53 |
| SOX2 | Sry (sex determining region y)-box 2 |
| TF | transcription factor |

| | |
|---|---|
| THL | National Institute for Health and Welfare (in Finnish Terveyden ja hyvinvoinnin laitos) |
| VPR | VP64-p65-Rta activation system |

# 1. Introduction

## 1.1. Lymphoblastoid cell lines

### 1.1.1. Epstein-Barr virus mediated transformation of B cells

Epstein-Barr virus (EBV) is a herpes virus, which usually causes an asymptotic infection, where it latently infects B cells, but it can also cause infectious mononucleosis, and the virus has been causally associated to multiple lymphomas and carcinomas, such as Burkitt's lymphoma and Hodgkin's disease (reviewed in Niedobitek et al. 2001). *In vitro* EBV infection of peripheral blood G0 phase B cells transforms the cells into actively proliferating lymphoblastoid cell lines (LCLs) (Sugimoto et al. 2004). The EBV genome resides in the infected cells as a high copy episome, which expresses a small subset of its genes (Babcock et al. 2000). The most upstream gene, which governs all the necessary functions of the EBV genome is the EBV nuclear antigen 1 (EBNA-1). It regulates the episome replication through interactions with the origin of plasmid replication (OriP) (Yates et al. 1985) and controls the faithful partitioning of the episome copies to daughter cells during mitosis by interacting with the segregating chromosomes (Wu et al. 2000). EBNA-1 is also the most upstream gene regulating the transformation process, as it activates the EBV transforming genes (Altmann et al. 2006), which subsequently activate the expression of several host cell genes related to cell cycle control and proliferation, such as cyclin D2, cyclin dependent kinase 4 and proto-oncogene C-MYC (Kaiser et al. 1999; Spender et al. 2001), resulting in the induction of proliferation.

The EBV-transformed LCLs can sometimes undergo immortalization, and the cell lines can be divided into two groups based on their immortalization status: the pre-immortal and the post-immortal LCLs. Pre-immortal LCLs are characterized by having a high proliferation rate, normal karyotype, and a low or non-existing telomerase activity (Sugimoto et al. 2004). These LCLs can be maintained up to 160 population doubling levels, after which they go into a proliferation crisis due to the shortening of telomeres, and the cell population dies (Sugimoto et al. 2004; Sugimoto et al. 1999). Post-immortal LCLs are also highly proliferative but, in contrast to pre-immortal LCLs, they are characterized by aneuploidy or other chromosomal aberrations as well as high telomerase

activity (Sugimoto et al. 2004), and they can be tumorigenic (Takahashi et al. 2003). Because of the high telomerase activity, the post-immortal cells do not undergo the proliferation crisis and can be maintained in culture seemingly infinitively (Sugimoto et al. 2004).

### 1.1.2. Research with biobanked LCLs

Biobanks and cell repositories have been using the establishment of LCLs as a convenient method to ensure a perpetual source of patient specific DNA and other biomolecules for over two decades. There are vast collections of LCLs from patients with various diseases, including rare genetic disorders, which offer a rich bioresource for disease research. For example, the NIMH Repository and Genomics Resource alone stores over 96,000 LCLs from patients with different mental disorders and control groups and almost the same number of DNA samples derived from LCLs, available for researchers (NIMH 2018). Some collections also include samples from non-affected family members or already established comprehensive genotypic or phenotypic datasets (Auton et al. 2015), further increasing their research value.

Biobanked LCLs have been widely used as a surrogate *in vitro* cell model when the primary tissue is not readily available. This is the case with disorders affecting the central nervous system, and LCLs have been used in a wide array of neurological and psychiatric disorder studies. These studies have unveiled for example new risk loci for schizophrenia (Ripke et al. 2011), depression (Holmans et al. 2007) and bipolar disorder (Sklar et al. 2011), as well as disease specific gene regulation patterns in autism (Hu et al. 2006). LCLs have also been used to study the host response to gene knockdowns (Mei et al. 2006; Badhai et al. 2009), drugs (Watters et al. 2004; Duan et al. 2007) and other cell perturbations (Correa & Cheung 2004; Niu et al. 2010).

Even though LCLs have been a conducive surrogate cell model in numerous studies, their ability to faithfully recapitulate the regulatory prospects of the primary tissue cells has been negated in several studies. The EBV transformation may cause differential expression levels of a subset of genes (Carter et al. 2002; Min et al. 2010; Çalişkan et al. 2011) or alter the DNA methylation patterns (Çalişkan et al. 2011; Hannula et al. 2001). LCLs have been shown to exhibit cell type specific alternative splicing (Monnier et al. 2003), and the chromosomal aberrations in post-immortal LCLs may affect research results (Redon et al. 2006; Okubo et al. 2001; Sugimoto et al. 2004). As a

consequence, the extensive LCL collections can be used only for certain type of studies, precluding their thorough utilization.

Recently, the emergence of iPSC technology has offered new perspectives for the use of the biobanked LCL collections. With the technology, the vast collections could be utilized more comprehensively in wider research applications.

## 1.2. Pluripotency reprogramming and induced pluripotent stem cell technology

### 1.2.1. Induced pluripotent stem cells

As cells differentiate from a pluripotent into a somatic state, they become gradually fixed on just one cell fate. Once fully matured, the cells have obtained a cell type specific epigenetic state, which governs how genes are regulated, thus dictating their phenotype. The epigenetic state of the differentiated cells is "locked" by DNA methylation patterns, which ensures that the cells stay in that particular state and cannot spontaneously revert back to pluripotency or to any other cell fate. However, somatic cells can be reprogrammed *in vitro* into a pluripotent state resembling that of an embryonic stem cell (ESC) with the ectopic overexpression of a small set of pluripotency-associated transcription factors (TFs) known as reprogramming factors. The most well-known factors are the so-called "Yamanaka factors": OCT4, SOX2, KLF4 and C-MYC (OSKM) (Takahashi & Yamanaka 2006; Takahashi et al. 2007). The reprogrammed cells are known as induced pluripotent stem cells (iPSCs). iPSCs are indistinguishable from ESCs, as they are pluripotent, have an unlimited self-renewal capacity and the potential to differentiate into all three germ layers, and they can contribute to the germline cells. Other pluripotent stem cell (PSC) characteristics are the high alkaline phosphatase level, which is often used for initial iPSC characterization; high nucleus-to-cytosol ratio and tightly packed colonies in cell culture; as well as the expression of pluripotency related markers, such as OCT4, SOX2, TRA-1-60 and TDGF1 (Martí et al. 2013).

### 1.2.2. Reprogramming to pluripotency

To reprogram somatic cells, reprogramming factors are introduced to the cells as transgenes, mRNAs or proteins (Hu 2014). One of the most used methods is the introduction of the factors as transgenes with non-integrative OriP/ EBNA-1 episomal plasmids (Okita et al. 2011). The ectopic overexpression of the reprogramming factors in cells results in the factors binding to any available target sites in the host genome (Chronis et al. 2017). This binding causes an array of stochastic events, which leads to a multistep process that can eventually result in the conversion of somatic cells into iPSCs. The complicated process is not well understood, but recent studies have provided some insights into the whole phenomenon (reviewed in for example Wang et al. 2017; Buganim et al. 2013).

The key event in reprogramming is the reconfiguration of the epigenome from the somatic into the pluripotent state. During the somatic state, the cell-type-specific genes are active, while the pluripotency-associated genes are silenced with such silencing histone markers as the histone H3 lysine 27 trimethylation (H3K27me3) (Wang et al. 2017). During the pluripotent state, on the other hand, the pluripotency-associated genes are active, and marked with histone markers such as histone H3 lysine 27 acetylation (H3K27ac) (Wang et al. 2017). The pluripotent state has also a characteristic poised state, where the different developmental pathways are suppressed but kept in a state, in which they are ready to respond to any differentiation cues (Wang et al. 2017). The reprogramming process can be divided into a stochastic and a hierarchal phase (Buganim et al. 2012). During the initial stochastic phase, the ectopic expression of reprogramming factors causes stochastic gene activation as they bind to their target sites. The binding causes the first major wave of transcriptomic (Polo et al. 2012), proteomic (Hansson et al. 2012) and epigenomic (Knaupp et al. 2017) reconfiguration, during which the cells start to lose their identity, undergo metabolic changes and gain higher proliferation rate (Polo et al. 2012; Hansson et al. 2012; Smith et al. 2010). The reprogramming factors can recruit chromatin remodeling complexes to their binding sites, and induce changes to the chromatin accessibility (Koche et al. 2011; Knaupp et al. 2017). Chromatin remodeling during the initial phase includes the silencing of the somatic gene regulatory elements (REs), that is the enhancers and promoters, and the activation of the early pluripotency gene REs (Knaupp et al. 2017). There is also a group of REs, which are transiently opened during the reprogramming process, but closed before the process is completed, possibly facilitating in the

chromatin accessibility during the process (Knaupp et al. 2017). The closing of active loci requires the replacement of the activating histone markers, such as the H3K27ac, with repressing ones, like H3K27me3, while during the activation of silenced loci, the markers are replaced vice versa. Throughout the early and intermediate phases of reprogramming, the cells gradually lose their cell-type-of-origin-specific DNA methylation patterns, mostly through passive, replication-dependent demethylation (Knaupp et al. 2017).

It has been suggested, that after the first wave the cells must undergo the stochastic activation of at least one of the genes, that belong to the subset of early pluripotency genes, i.e. Esrrb, Utf1, Lin28 and Dppa2, which can initiate the activation of the rest of the pluripotency network (Buganim et al. 2013). The activation of these early pluripotency genes is thought to be the rate-limiting step during the reprogramming process, as most of the cell population goes through the initial stochastic events but only a fraction of the cells can give rise to iPSCs. Once the rate-limiting step is passed, the early pluripotency genes activate SOX2, which is considered to mark the onset of the late hierarchical phase (Buganim et al. 2012). The endogenous SOX2 governs the activation of the rest of the pluripotency network, and the second major wave of reconfiguration occurs in the cells (Buganim et al. 2012). The somatic loci are silenced, and the late pluripotency-associated loci are activated as they become more accessible for the chromatin remodeling complexes, thus constituting an ESC-like epigenomic state (Knaupp et al. 2017). During the late phase, the DNA methylation starts to occur more rapidly and seems to rely on the higher expression of the endogenous demethylases TET1 and TET2 (Knaupp et al. 2017). The genome also acquires new methylation patterns, and it seems that once the ESC-like epigenome has been established, the new DNA methylation patterns “lock” the state (Knaupp et al. 2017). Once the pluripotent state has been acquired, the ectopic reprogramming factor genes are silenced, and the pluripotent state is stabilized (Golipour et al. 2012).

#### 1.2.2.1. Reprogramming of different cell types

Most of the studies about the reprogramming process have been conducted with human or murine fibroblasts. It was recently demonstrated that depending on the cell type of origin the cells undergo cell-type-specific transcriptomic changes and the reprogramming process follows a cell-type-

specific trajectory (Nefzger et al. 2017). Since different cell types express different sets of genes, which make up their phenotype, the loss of identity occurs in a cell-type-specific manner, as the cell identity related genes are downregulated (Nefzger et al. 2017). The re-activation of the pluripotency network occurs also partially in a cell-type-specific manner, and it was discovered that for example the mesenchymal-to-epithelial transition (MET), which has been considered crucial for reprogramming based on studies in fibroblasts, does not occur in keratinocytes (Nefzger et al. 2017). The transcriptome analysis nevertheless revealed some universal events that took place in all studied cell types. All cell types exhibited two major waves of transcriptomic resetting in the beginning and end of the reprogramming process, though the genes that were differentially expressed during the waves in each cell type varied (Nefzger et al. 2017). There were also some somatic genes, which were universally downregulated during the loss of identity, such as Egr1, which has been shown to impede reprogramming (Nefzger et al. 2017). The *bona fide* iPSCs generated from different cell types of origin were indistinguishable based on transcriptome studies (Nefzger et al. 2017), though other studies have shown that early passage iPSCs differ from each other based on their cell type of origin in their epigenetic memory, but the differences are abolished with further passaging (Polo et al. 2010).

### 1.2.3. Reprogramming factors

Reprogramming has been widely conducted and studied by using the four Yamanaka factors, but additional factors have been discovered to improve the reprogramming efficiency or to substitute some or all of the canonical factors (reviewed in Xiao et al. 2016; Theunissen & Jaenisch 2014). The reprogramming factors are usually pluripotency-associated TFs, that exhibit master TF as well as pioneer factor properties. As a master TF, the reprogramming factors are capable of controlling the expression of a set of genes required for a specific cell identity program, in this case the pluripotency program. Pioneer factors, on the other hand, are capable of recognizing their target loci in closed chromatin regions and making the loci more accessible for other TFs to bind. This is necessary, since during the somatic cell program, the pluripotency related loci reside in closed chromatin. In this thesis work, the effects of eight reprogramming factors were studied: OCT4, SOX2, KLF4, C-MYC, L-MYC, LIN28, NANOG and REX1.

OCT4

OCT4 (octamer-binding transcription factor 4, also known as POU5F1) was the first master regulator of the pluripotent state to be discovered (Nichols et al. 1998). It is one of the core pluripotency factors together with SOX2 and NANOG, and it regulates several pluripotency-associated functions in the cell, such as proliferation, stem cell maintenance and cell metabolism (Kim et al. 2008; Xiao et al. 2016). OCT4 could be considered the most important pluripotency factor and it can be replaced in the reprogramming factor combinations only with factors that are able to activate the endogenous OCT4 (Radzisheuskaya & Silva 2014; Xiao et al. 2016). During reprogramming, OCT4 forms a heterodimer with SOX2, and the dimer functions as a pioneer TF, which is capable of activating pluripotency genes (Tapia et al. 2015). OCT4 has an important role in chromatin remodeling as it has been shown to interact with several chromatin remodeling complexes (van den Berg et al. 2010). OCT4 has also been suggested to facilitate MET (Radzisheuskaya & Silva 2014).

SOX2

SOX2 (sex determining region of the Y-box 2) is one of the core regulators of the endogenous pluripotency network (Kim et al. 2008). In ESCs, SOX2 has an important function in maintaining self-renewal and pluripotency (Xiao et al. 2016). The endogenous SOX2 plays an important role during the reprogramming process as its activation has been hypothesized to mark the onset of the hierarchal phase (Buganim et al. 2012). During reprogramming the ectopic SOX2 functions as a OCT4-SOX2 heterodimer TF (Tapia et al. 2015).

KLF4

KLF4 (Krüppel-like factor 4) is related to regulation of proliferation, apoptosis, tissue homeostasis, embryo development and cytodifferentiation (Xiao et al. 2016). It is irreplaceable in naïve human iPSC state, where it controls the naïve pluripotency circuitry (Liu et al. 2017). During reprogramming, it is crucial to express the factor at a high enough level for the process to be completed correctly, and arrested reprogramming can be saved in some cases with upregulation of KLF4 (Nishimura et al. 2014). KLF4 seems to have a dual function during the reprogramming process, since during the early phase it has been shown to repress genes related to differentiation (Polo et al. 2012) and

during the late phase it activates pluripotency genes by promoting transcriptional pause release (Liu et al. 2014).

### C-MYC and L-MYC

C-MYC and L-MYC are members of the MYC family of TFs. C-MYC functions as general amplifier of transcription by inducing the RNA polymerase II pause release (Nie et al. 2012; Rahl et al. 2010). It is involved in several important functions in PSCs, such as cell growth and proliferation, metabolism, differentiation, apoptosis, senescence, stem cell renewal, embryonic development and cell adhesion (Xiao et al. 2016). During reprogramming it amplifies the expression of all active genes, both pluripotent and somatic ones (Nie et al. 2012). Reprogramming is possible without C-MYC, but it greatly enhances the efficiency (Nakagawa et al. 2008). In iPSCs, the transcriptional network governed by C-MYC has been shown to be separate from the core pluripotency network (Kim et al. 2008). However, C-MYC is also a known oncogene, but it can be readily replaced with L-MYC, which improves the reprogramming efficiency of human fibroblasts more than C-MYC does (Nakagawa et al. 2008). L-MYC is similar to C-MYC in function, except it lacks the domain which contributes to the oncogenic properties, making it a safer choice for iPSC generation (Okita & Yamanaka 2011).

### NANOG

NANOG is the third main factor in the core pluripotency network, together with OCT4 and SOX2 (Kim et al. 2008), and required for the formation and maintenance of pluripotency in ESCs (Mitsui et al. 2003). Nanog is a very specific marker of pluripotent cells and has been shown to contribute to the ground pluripotency state (Silva et al. 2009). However, NANOG is not required for the induction of pluripotency, but it improves the reprogramming efficiency or can be used to replace other factors (Schwarz et al. 2014; Xiao et al. 2016).

### LIN28

LIN28 (also known as LIN28A) has been considered to be a robust marker during the intermediate phase of reprogramming, marking the cells that will undergo the full reprogramming process (Buganim et al. 2013). It has been suggested that LIN28 is one of the genes that can either directly

or through other factors activate the endogenous SOX2 (Buganim et al. 2012). In PSCs, LIN28 inhibits differentiation by repressing Let-7 microRNAs (Viswanathan & Daley 2010). During reprogramming, it also recruits the TET1 chromatin remodellers to closed genomic sites to make them more accessible (Zeng et al. 2016).

### REX1

REX1 (also known as ZFP42) is expressed in undifferentiated cells (Chen & Gudas 1996). It is directly regulated by OCT4 and indirectly regulated by SOX2 and NANOG, making it an important part of the pluripotency network (Ben-Shushan et al. 1998; Shi et al. 2006). It has a role in development, maintaining the undifferentiated state as well as cell cycle regulation (Scotland et al. 2008). REX1 has also been shown to improve reprogramming by impeding such reprogramming barriers as growth arrest and apoptosis (Son et al. 2013).

### Other factors

On top of the transcription factors used to reprogram cells, there are several other factors, which are used to improve reprogramming or to replace reprogramming factors. These factors can be small molecules or micro-RNAs (miRNAs) (reviewed in Xiao et al. 2016). Also, endogenous factors or signaling pathways, which impede the reprogramming process, can be repressed with interference methods to improve reprogramming. In this thesis work, sodium butyrate (NaB) and short hairpin RNA against p53 (shp53) were used. NaB improves the reprogramming process by upregulating the expression of pluripotency-related miR302/367 cluster and increasing the stability of the miRNAs (Zhang & Wu 2013). This miRNA cluster has been shown to promote MET and thus enhance reprogramming (Liao et al. 2011). Reprogramming has been shown to induce a cellular stress response, which causes the upregulation of p53, which in turn can induce either apoptosis or cell cycle arrest and therefore interfere with the reprogramming process (Marión et al. 2009). Expression of shp53 causes the repression of p53, decreases the stress response and improves the reprogramming efficiency (Hong et al. 2009).

### 1.2.4. Prospects of iPSC technology

iPSC technology holds great promise for the development of disease models, drug discovery and regenerative medicine, and it has resolved several ethical issues surrounding the use of ESCs in similar research (Shi et al. 2017). In disease modeling, virtually any cell type from a patient can be converted into iPSCs, which are then differentiated into the cell type or tissue affected by the disease and the effects of the disease can be studied *in vitro*. Monogenic diseases have been widely studied since the disease-causing mutation can be corrected with a gene editing platform, such as the CRISPR/Cas9, which results in an isogenic control. Similarly, a mutation can be introduced into a healthy iPSC line to study its effects. With an isogenic control, the only differences between the healthy and disease cells are the ones caused by the disease mutation. The iPSC technology enables also the study of disease-specific cells which would be unattainable otherwise, such as neurons, and for example several iPSC-derived neuronal models have been generated for Alzheimer's disease alone (reviewed in Shi et al. 2017). iPSC-derived somatic cells offer also a new strategy for drug discovery and research (Shi et al. 2017). Drug testing has been heavily dependent on animal models, but there are inadequacies regarding the capability of animals to recapitulate all the effects of the diseases or tested drugs that occur in humans (Puzzo et al. 2015). With iPSC technology, the effects of drugs on their target cell types, such as neurons, can be studied *in vitro*, which can reduce the amount of expensive, failed late-stage clinical trials in the future (Shi et al. 2017). Wide drug screens have also been conducted to study the effects of already existing drugs on other cell types (Yamashita et al. 2014). In iPSC technology derived regenerative medicine, patient derived iPSCs are generated, possible mutations corrected, and the cells are then differentiated into the desired cell type. These somatic cells are used to replace damaged or non-functioning tissue. iPSCs offer the possibility to generate autologous, non-immunogenic somatic cells for patients, thus mitigating the potential health and safety issues, which could rise from allologous ESC-derived cells (Shi et al. 2017). Clinical trials using iPSCs in regenerative medicine have been focused mostly on eye diseases and neural repair (reviewed in Takahashi 2018).

iPSC technology also offers new possibilities for wider utilization of the LCL collections since the cells could be reprogrammed and differentiated into the desired cell type to study the disease specific cell types *in vitro*. By utilizing the already existing collections, researchers would not have to collect new samples from patients, thus saving time and resources. LCLs have been

reprogrammed previously into bona fide iPSCs (Kumar et al. 2016; Choi et al. 2011; Rajesh et al. 2011; Barrett et al. 2014; Thomas et al. 2015; Banovich et al. 2018; Son et al. 2017). Given that most of the reprogramming methods have been developed for fibroblasts, LCLs have been proven to need optimization of the conventional methods in order to be efficiently reprogrammed (Kumar et al. 2016). iPSC technology derived research has been modest with LCLs, but LCL-derived iPSCs have been used to model Parkinson's disease (Fujimori et al. 2016). In the study, iPSCs derived from the LCL and fibroblasts of a Parkinson's disease patient were generated and differentiated into neural cells. The iPSCs from both cell types were found to have the same differentiation potential, and no differences were found between the derived neural cells. These results demonstrated that LCLs hold great potential for disease modeling and other iPSC technology related research.

In order to fully utilize the potential of the LCL collections, more efficient reprogramming methods should be developed which would result in higher quality iPSCs. To achieve this, the endogenous events, that are crucial for successful reprogramming of LCLs, have to be understood. However, the conventional reprogramming methods rely mostly on the overexpression of ectopic reprogramming factors, which instigate an array of stochastic events. The abundance of these events masks the reprogramming-wise necessary ones. This makes it impossible to identify the key factors or epigenomic changes, that should be targeted during reprogramming to achieve more efficient process and higher quality iPSCs. In order to systematically study the endogenous reprogramming events, the CRISPR/Cas genome editing platform has been harnessed for reprogramming in fibroblasts but is yet to be established in LCLs.

## 1.3. CRISPRa mediated reprogramming

### 1.3.1. CRISPR/Cas

Clustered regularly interspaced palindromic repeats (CRISPR)/CRISPR-associated (Cas) system has been adapted for gene editing purposes from the acquired immunity of prokaryotes (Jinek et al. 2012; Mali, Yang, et al. 2013). The technology has enabled a fast, easy and inexpensive way to create specific gene mutations. It is composed of only two components: the Cas9 protein and a guide-RNA (gRNA), which are easily engineered and expressed in cells to achieve the desired mutation (Jinek

et al. 2012). Cas9 is an endonuclease which has two nuclease domains, making it capable of eliciting double stranded breaks (DSBs) to DNA (Garneau et al. 2010). The gRNA is a short RNA fragment, which has a Cas9 binding region and a spacer region (Jinek et al. 2012). Once both components are expressed in cells, the gRNA binds to Cas9 and the complex is then guided to its target site in the host genome, based on the ~20 nucleotides long spacer sequence, which is complementary to that in the target site. The target site in the host genome is called a protospacer, and it has to contain a protospacer adjacent motif (PAM) after its 3' end, for the Cas9 to be able to bind to it. Different prokaryotic species have different PAM sequences for their Cas9 proteins (Mojica et al. 2009). The most widely used Cas9 is from *Streptococcus pyogenes*, the PAM sequence of which is NGG, N being any nucleotide (Mojica et al. 2009). Once the Cas9-gRNA complex has bound to its target site, the Cas9 undergoes a conformational change so that it is able to elicit a DSB near the target site (Jinek et al. 2014). The DSB is repaired by the repair machinery of the cell, either via non-homologous end joining, which usually results in knock-out mutants, or homology directed repair, via which it is possible to introduce new genetic material into the cells (Mali, Yang, et al. 2013).

### 1.3.2. CRISPR/Cas9 mediated gene activation

By introducing point mutations to the two active sites of the Cas9 endonuclease, it is possible to render the protein catalytically inactive without interfering with its DNA binding properties. This deactivated, "dead" Cas9 (dCas9), is a protein which is capable of binding to essentially any genomic site with a PAM sequence next to it, guided by a short gRNA, and several applications of the platform have been developed by engineering either one or both of the components. These applications enable for example labeling of sequence specific loci in the genome (Knight et al. 2018), remodeling of chromatin (Pulecio et al. 2017) and multiplexed regulation of gene circuitries (Jusiak et al. 2016). In this thesis the CRISPR/dCas9 platform was used for gene activation.

In CRISPR mediated gene activation (CRISPRa), a transcriptional activator domain is fused to the dCas9 and the protein complex can be used as an artificial TF when directed to the promoter region of a target gene (Perez-Pinera et al. 2013; Maeder et al. 2013). The dCas9 complex can functionally open chromatin when targeted to closed chromatin regions, thus exhibiting pioneer factor properties (Barkal et al. 2016). Various transcriptional activator domains, which have

different mechanisms of function, have been used for CRISPRa (table 1). Some activator domains recruit the transcriptional machinery to the target site (e.g. VP64), while others activate gene expression by marking the histones with activating H3K27ac markers (p300). More advanced activator systems constitute of several domains synergistically improving the gene activation, and for example the SunTag system can recruit multiple activator domains to a single dCas9 protein (Tanenbaum et al. 2014).

**Table 1. Different activator domains and systems used for CRISPRa**

| Activator domain/ system | Components | Targeted locus | Mechanism of function | References |
|---|---|---|---|---|
| VP64 | 4 tandem repeats of the viral VP16 transcriptional activator | Promoters | Recruitment of transcriptional activators | (Perez-Pinera et al. 2013; Maeder et al. 2013) |
| VP160 | 10 repeats of VP16 | Promoters | Recruitment of transcriptional activators | (Cheng et al. 2013) |
| VP192 | 12 repeats of VP64 | Promoters | Recruitment of transcriptional activators | (Balboa et al. 2015) |
| p65 | The NF-kB trans-activating subunit | Promoters | Recruitment of transcriptional activators and chromatin remodellers | (Gilbert et al. 2013; Konermann et al. 2015) |
| p300 core | The catalytic core of acetyltransferase | Promoters and enhancers | Histone acetyltransferase | (Hilton et al. 2015) |
| VPR | VP64-p65-Rta | Promoters | Recruitment of transcriptional activators | (Chavez et al. 2015) |
| SunTag | Multiple effector domains (such as VP64 or p65) recruited to single peptide fused to dCas9 | Depends on effector domains | Depends on effector domains | (Tanenbaum et al. 2014; Gilbert et al. 2014) |
| SAM | VP64 fused to dCas9, gRNA with binding sites for two MS2 aptamers, MS2-p65-HSF1 | Promoters | Recruitment of transcriptional activators | (Konermann et al. 2015) |

Because the gRNAs are short and easily expressed in cells, it is possible to simultaneously target and regulate several genes or complete gene networks in cells (Cheng et al. 2013; Dominguez et al. 2015). This high multiplexing capability of the CRISPR/dCas9 platform and the pioneer TF properties of the dCas9 offer the possibility to activate the endogenous pluripotency network genes and induce reprogramming without any ectopic reprogramming factors.

### 1.3.3. CRISPRa mediated reprogramming

CRISPRa mediated reprogramming has already been demonstrated in mouse and human fibroblasts. The mouse work focused on studying the necessary endogenous events required for reprogramming (Liu et al. 2018). In their work, Ding and colleagues first showed that iPSCs can be generated by targeting several factors of the pluripotency network with a dCas9-SunTag-VP64 system (Liu et al. 2018). They also demonstrated that by using a dCas9-SunTag-p300core system, the targeting and epigenetic remodeling of the promoter region of *Sox2* or the promoter and enhancer regions of *Oct4* are enough to induce successful reprogramming of mouse fibroblasts (Liu et al. 2018). The work demonstrated the importance of epigenetic remodeling during reprogramming, and also that the effects of individual events on reprogramming can be studied systematically with CRISPRa without the masking effects of the ectopic factors (Liu et al. 2018).

In the human fibroblast reprogramming, Otonkoski and colleagues (Weltner et al. 2018) showed that the CRISPRa system enables the utilization of genomic elements to improve the reprogramming process, which is impossible with transgenic methods. They targeted the promoter regions of several pluripotency network genes as well as the recently discovered EEA motifs with a dCas9-VP192 activator complex (Weltner et al. 2018). The EEA motif, or embryonic genome activation (EGA) enriched Alu element motif, is a 36-base-pair-long genomic element that is enriched in the promoter regions of genes related to EGA (Töhönen et al. 2015). The EEA motif is abundant in the human genome and it has been suggested to have a role during EGA (Töhönen et al. 2015). In CRISPRa mediated reprogramming, the targeting of the EEA motif was shown to improve the reprogramming efficiency by an order of magnitude (Weltner et al. 2018). The targeting of the EEA motif enhanced the expression of a subset of pluripotency genes, including NANOG and REX1 (Weltner et al. 2018). According to the suggested mechanistic model, the dCas9 complex binding to the EEA motifs in the promoters of pluripotency-associated genes renders the chromatin more accessible near the binding sites, thus facilitating the activation of the genes (Weltner et al. 2018).

The novel CRISPRa mediated reprogramming method could enable us to systematically determine the necessary endogenous events required for efficient reprogramming of LCLs and to utilize the genomic elements in the process. With the method, the speculated rate-limiting steps, such as the stochastic activation of *Esrrb*, *Utf1*, *Lin28* or *Dppa2*, could be directly

targeted and their effects could be studied. This will help in the development of more efficient and cell-type-specific reprogramming methods. By specifically targeting the crucial factors and events during reprogramming, the resulting iPSCs could also prove to be of higher and more standardized quality, which subsequently results in higher-quality downstream applications.

# 2. Aim of the thesis

The emergence of iPSC technology has offered new possibilities for the more thorough utilization of the vast lymphoblastoid cell line collections in biobanks worldwide. In order to develop more efficient reprogramming methods, which will give rise to higher-quality iPSCs, it is important to understand the cell-type-specific endogenous events taking place during the reprogramming process. The novel CRISPRa mediated reprogramming method enables the systematic study of the endogenous reprogramming events without the masking effects of the ectopic factors and stochasticity in transgenic reprogramming methods. The CRISPRa method also enables the utilization of genomic elements, which is not plausible with conventional methods.

This thesis work aimed to establish the novel CRISPRa mediated reprogramming method, described by Otonkoski and colleagues (Weltner et al. 2018), for the first time in biobanked LCLs. The aim was to derive bona fide iPSC lines from the LCLs, obtained from the Psychiatric Family Collections of the THL Biobank, characterize the generated lines and return them back to the biobank as a future research resource. The generated collection would serve as the first ever iPSC line collection derived from biobanked cells with the CRISPRa method and would pave the way for the wider use of the collected LCLs.

# 3. Materials and methods

## 3.1. Lymphoblastoid cell lines

The lymphoblastoid cell lines used in this project were obtained from THL Biobank by an application process. LCLs were selected from the Psychiatric Family Collections of THL Biobank and selection was made in collaboration with the original researchers based on the future research potential of produced iPSCs in biobank. The lines were labeled as IB-D1 through IB-D6 in this project. Experimental work with the cell lines was carried out under the contract and general terms of THL Biobank (permission no BB2016_56).

## 3.2. Plasmids

The activator plasmid was pCXLE-dCas9VP192-GFP-shp53 (Addgene #69535) and gRNAs were cloned to GG-dest plasmid (Addgene #69538). The gRNAs used in this thesis were OCT4.1, OCT4.4, OCT4.5, SOX2.4, SOX2.3, KLF4.2, KLF4.3, CMYC.3, CMYC.5, EEA.1, EEA.2, EEA.3, EEA.7, EEA.10, NANOG.3, NANOG.4, NANOG.1, NANOG.2, NANOG.5, LIN28.1, LIN28.2, LIN28.3, LIN28.4, LIN28.5 REX1.10 and REX1.8 as described in (Weltner et al. 2018). Transgenic reprogramming plasmids were pCXLE-hOCT3/4-shp53-F (Addgene #27077), pCXLE-hSK (Addgene #27078), and pCXLE-hUL (Addgene #27080). Control GFP plasmid was pCXLE-EGFP (Addgene #27082).

Plasmids were prepared from glycerol stocks with NucleoBond® Xtra Midi Plus EF kit (Macherey-Nagel, 740422) according to the user manual, yielding a final concentration of 0.5 – 2 µg/µl, measured with SimpliNano™ Spectrophotometer (Biochrom).

## 3.3. Cell culture

Lymphoblastoid cells were cultured in lymphoblastoid cell medium (RPMI 1640 with GlutaMAX™ (Life Technologies, 61870-010), 15 % fetal bovine serum (FBS; Life Technologies, 10270-106), 1 µM

Sodium Puryvate (Lonza, BE13-115E)) in suspension, in cell culture flasks which were kept upright. Cell density was kept between 0.5 – 1.5 * $10^6$ cells/ml. Culture was passaged as follows: cells were transferred to a conical tube, centrifuged 200 g, 5 min, at room temperature (RT) (Eppendorf, Centrifuge 5810 R), and the cell pellet was re-suspended to new medium. When transferring cells to conical tube and re-suspending the cell pellet, the culture was pipetted multiple times to ensure single cell suspension. Cells were counted from a single cell suspension sample taken from transferred culture with Countess® Automated Cell Counter (Life Technologies) according to the user manual. New medium was added when needed between passaging without disrupting cell clusters.

Human foreskin fibroblasts (HFFs; ATCC line CRL-2429) were cultured in fibroblast medium (DMEM (Sigma, D6546), 10 % FBS (Life Technologies, 10270-106), 1 % GlutaMAX™ (Life Technologies, 35050-038)). Culture was passaged once the confluence was above 80 % as follows: cells were washed twice with phosphate buffered saline (PBS; Sigma, D8537), detached with TrypLE™ (Life Technologies, 12563-029) by incubating at +37 °C, 3 min, fibroblast medium was added and the culture was centrifuged 5 min, 200 g, RT. The cell pellet was re-suspended to fibroblast medium, cells were counted with Countess®, and plated with desired dilution.

iPSCs were cultured on Matrigel (MG; Corning, 356231) coated plates in Essential 8 (E8) medium (Life Technologies, A1517001) or in E8 Flex (Life Technologies, A2858501). E8 was changed every day, and E8 Flex enabled medium change intervals of up to 72 hours. iPSCs were passaged when colonies had grown to optimal size to remain viable after passaging. Passaging was done as follows: medium was aspirated, 0.5 mM EDTA (Life Technologies, 15575020) was added and incubated at RT until colonies started to detach slightly from the plate, EDTA was aspirated, new medium was added, and cells were scraped from the plate with a pipet tip, after which the cells were transferred to a new plate in a desired dilution.

All cells were kept in Forma Steri-Cycle $CO_2$ Incubator (Thermo Scientific) at +37 °C with 5 % $CO_2$.

## 3.4. Transfection

Cells were transfected by electroporation using Neon™ Transfection System (Invitrogen) according to the user manual. When electroporating lymphoblastoid cells, the culture was split to 0.5 * $10^6$ cells/ml density on the previous day to obtain optimal cell growth phase. 2 * $10^6$ cells and 6 µg of total DNA were used for one electroporation with a 100 µl tip. Electroporation conditions for lymphoblastoid cells were 1300 V, 10 ms and 3 pulses using the R buffer from the Neon™ Transfection System Kit (Invitrogen, MPK10096).

Fibroblasts were detached from the plate with TrypLE™ treatment as in passaging. 1 * $10^6$ cells and 6 µg of total DNA were used for one electroporation with a 100 µl tip. Electroporation conditions were 1650 V, 10 ms, 3 pulses with R buffer.

## 3.5. FACS

Fluorescence-activated cell sorting (FACS) samples were prepared as follows: 1 ml of cell suspension was transferred to a microcentrifuge tube, centrifuged 200 g, 5 min, at RT (LaboGene, ScanSpeed Mini), and the cell pellet was re-suspended to 500 µl of FACS-buffer (PBS (Sigma, D8537), 5 % FBS (Life Technologies, 10270-106)). FACS samples were filtered through a 30 µm filter to obtain a single cell suspension. Samples were run with BD FACSCalibur™ (BD Biosciences) and results analyzed with FlowJo® software.

## 3.6. Reprogramming and iPSC line propagation

To reprogram LCLs, the cells were transfected with reprogramming plasmid combinations. After transfection, cells were put to 6-well plates with hES + NaB medium (KnockOut DMEM (Life Technologies, 10829018), 20% KO serum replacement (Life Technologies, 10828028), 1% GlutaMAX™ (Life Technologies, 35050-038), 0.1 mM β-mercaptoethanol (Life Technologies, 31350-010), 1% nonessential amino acids (Life Technologies, 11140050), 6 ng/ml basic fibroblast growth

factor (bFGF; Sigma, SRP4037), 0.25 mM sodium butyrate (NaB; Sigma-Aldrich, B5887)) for expansion, before transferring them to 10 cm MG plates for seeding. Cells were grown in hES + NaB medium until first iPSC colonies appeared, at which point the medium was changed to E8. Medium was changed every other day, first by centrifugation similarly to LCL culture, and once enough cells were attached to the plate, old medium was aspirated along with non-attached cells, and new medium was added. Cells were grown until the formed iPSC colonies were big enough for either alkaline phosphatase staining or iPSC line propagation, in any case up to maximum of 33 days.

For iPSC line propagation, iPSC colonies were manually picked by cutting a colony to smaller pieces with a scalpel (Feather, M03E/5203012) and transferring the colony fragments with a pipet to MG coated 24-well plates in E8 medium. Clones from one LCL were picked from different iPSC colonies. From a 24-well plate, clones were transferred to 6-well plates and cultured as iPSCs. Clones were passaged up to 10 times to obtain enough cells for iPSC line characterization samples, before freezing two vials of each clone.

## 3.7. Alkaline Phosphatase Staining

Alkaline phosphatase (AP) staining was performed as follows: reprogramming culture plates were fixed with 4 % paraformaldehyde (PFA; Fisher Chemical, 10131580) for 10 min, RT. Plates were washed three times with PBS, and stained with NBT/BCIP containing solution (200 µl NBT/BCIP Stock Solution (Roche, 11681451001) in 10 ml 0.1 M Tris-HCl pH 9.5, 0.1 M NaCl, 0.05 M $MgCl_2$) until AP activity was detected as precipitate development. Plates were washed with PBS to stop the reaction and scanned with Epson Perfection V500 Photo -scanner with 600 dpi resolution. The number of AP positive colonies was calculated manually from scanned images. Reprogramming efficiency was calculated as follows: $\frac{AP\ positive\ colonies}{Transfected\ cells} * 100\ \%$.

## 3.8. Immunocytochemistry

Cell culture samples for immunocytochemistry (ICC) were fixed with 4 % PFA for 30 min, RT. Cells were washed three times with PBS, and then permeabilized with PBS containing 0.2 % Triton-X 100

(Fisher Scientific, BP151-100) for 15 min, RT. Cells were washed again three times with PBS, and blocked with Ultra Vision protein block (Thermo Scientific, TA-125-PBQ) 10 min, RT. Blocking solution was aspirated and primary antibody mixture added. All antibodies were diluted in PBS with 0.1 % Tween20 (VWR International, 663684B). Primary antibodies were incubated for 48 h when staining for pluripotency markers, or for 24 h when staining for germ layer markers, at +6 °C on a See-saw rocker (Stuart, SSL4) with the given dilutions. After primary antibody incubation, cells were washed three times with PBS and secondary antibody mixture was added. Secondary antibodies were incubated for 30 min, RT, on a See-saw rocker. Secondary antibody mixtures contained 10 ng/µl Hoechst33342 (Sigma-Aldrich, B2261) to stain the nuclei. Used antibodies and dilutions in tables 2 and 3. Cells were imaged with EVOS FL microscope (Life Technologies).

Table 2. Primary antibodies for pluripotency staining

| Antibody | Host species | Dilution | Manufacturer | Product code |
|---|---|---|---|---|
| Oct3/4 | Rabbit | 1:500 | Santa Cruz Biotechnology | SC-9081 |
| Sox2 | Rabbit | 1:500 | Cell Signaling | 3579 |
| Lin28 | Rabbit | 1:500 | Cell Signaling | 8641 |
| Tra-1-60 | Mouse | 1:200 | Thermo Scientific | MA1-023 |
| Tra-1-81 | Mouse | 1:250 | Thermo Scientific | MA1-024 |
| Sox17 | Goat | 1:400 | R&D Systems | AF1924 |
| $\alpha$-smooth muscle actin | Mouse | 1:400 | Sigma | A2547 |
| $\beta$-tubulin III | Rabbit | 1:500 | Abcam | Ab18207 |

Table 3. Secondary antibodies

| Antibody | Host species | Dilution | Manufacturer | Product code |
|---|---|---|---|---|
| Anti-Goat IgG (H+L) Alexa Fluor 488 | Donkey | 1:500 | Invitrogen | A-11055 |
| Anti-Mouse IgG (H+L) Alexa Fluor 488 | Donkey | 1:500 | Invitrogen | A-21202 |
| Anti-Mouse IgG (H+L) Alexa Fluor 594 | Donkey | 1:500 | Invitrogen | A-21203 |
| Anti-Rabbit IgG (H+L) Alexa Fluor 488 | Donkey | 1:500 | Invitrogen | A-21206 |
| Anti-Rabbit IgG (H+L) Alexa Fluor 594 | Donkey | 1:500 | Invitrogen | A-21207 |

## 3.9. RT-qPCR

RNA samples for reverse transcriptase quantitative polymerase chain reaction (RT-qPCR) were collected as follows: iPSCs and fibroblasts were detached from the plate by scraping with a pipet tip, transferred to a microcentrifuge tube, centrifuged at 200 g, 5 min, RT, and medium was carefully aspirated. Lymphoblastoid cells were just transferred to microcentrifuge tubes and treated similarly from there-on. Cell pellets were re-suspended to 350 µl of LBP lysis buffer from NucleoSpin® RNA Plus kit (Macherey-Nagel, 740984). RNA was purified with the NucleoSpin® RNA Plus kit, and the concentration was measured with SimpliNano™ Spectrophotometer. RNA samples were concentrated when necessary with the Eppendorf Concentrator 5301.

800 ng of total RNA was used for one reverse transcriptase (RTase) reaction in 11.3 µl of Nuclease free $H_2O$, AccuGENE™ Molecular Biology water (DEPC water; Lonza, 51200), which was heated at +65 °C for 1 min to get rid of any secondary structure. RTase reaction master mix was prepared per one reaction as follows:

| | |
|---|---|
| 4 µl | M-MLV Reverse Transcription 5x buffer (Promega, M531A) |
| 2.5 µl | 2.5 mM dNTPs (Promega, U1420) |
| 1 µl | Oligo$(dT)_{18}$ primer (Thermo Scientific, SO132) |
| 0.2 µl | Random primers (Promega, C118A) |
| 0.5 µl | RiboLock RNase inhibitor (Thermo Scientific, EO0381) |
| 0.5 µl | M-MLV Reverse Transcriptase (Promega, M170A) |

8.7 µl of master mix was added to 11.3 µl RNA + DEPC water. Mixtures were vortexed well and spun down, before incubation at +37 °C, 90 min. The RTase was inactivated by incubating at +95 °C, 5 min, and 20 µl of DEPC water was added. A water control was used to ensure contamination-free reagents.

For qPCR reactions, a master mix containing 5x HOT FIREPol EvaGreen qPCR Mix Plus (Solis BioDyne, 08-25-0020) and DEPC water was prepared and divided to PCR strips containing the cDNA. 2 uM forward and reverse primer mixes (table 4) were prepared. QiAgility pipetting robot (Qiagen) pipetted the reactions of 20 ul in duplicates as follows:

| | |
|---|---|
| 4 ul | 5x HOT FIREPol EvaGreen qPCR Mix Plus |
| 5 ul | 2 uM F/R primer mix |
| 1 ul | cDNA |
| 10 ul | DEPC water |

qPCR was run with Qiagen Rotor-Gene Q or Rotor-Gene 6000 machine with the following program:

| | | | |
|---|---|---|---|
| Polymerase activation, initial denaturation | | 95 °C | 15 min |
| 40 cycles: | Denaturation | 95 °C | 25 sec |
| | Annealing | 57 °C | 25 sec |
| | Elongation | 72 °C | 25 sec |

Table 4. Primers for qPCR

| Gene | Origin | Forward primer[1] | Reverse primer[1] | Product length |
|---|---|---|---|---|
| OCT4 | human (NM_002701) | 5'-TTG GGC TCG AGA AGG ATG TG-3' | 5'-GTG AAG TGA GGG CTC CCA TA-3' | 193 bp |
| SOX2 | human (NM_003106) | 5'-GCC CTG CAG TAC AAC TCC AT-3' | 5'-TGC CCT GCT GCG AGT AGG A-3' | 85 bp |
| NANOG | human (NM_024865.2) | F 5'-CTC AGC CTC CAG CAG ATG C-3' | 5'-TAG ATT TCA TTC TCT GGT TCT GG-3' | 94 bp |
| REX1 | human (NM_174900.3) | 5'-CGT TTC GTG TGT CCC TTT CAA-3' | 5'-CCT CTT GTT CAT TCT TGT TCG T-3' | 106 bp |
| TDGF1 | human (NM_003212.2) | 5'-TCA GAG ATG ACA GCA TTT GGC-3' | 5'-TTC AGG CAG CAG GTT CTG TTT A-3' | 118 bp |
| Cyclophilin G | human (NM_004792) | 5'-CAA TGG CCA ACA GAG GGA AG-3' | 5'-CCA AAA ACA ACA TGA TGC CCA-3' | 94 bp |
| KLF4 | human (NM_004235.4) | 5'-CCG CTC CAT TAC CAA G-3' | 5'-CAC GAT CGT CTT CCC CTC TT-3' | 80 bp |
| CMYC | human (NM_002467) | 5'-AGC GAC TCT GAG GAG GAA CA-3' | 5'-CTC TGA CCT TTT GCC AGG AG-3' | 87 bp |
| LIN28 | human (NM_024674) | 5'-AGG AGA CAG GTG CTA CAA CTG-3' | 5'-TCT TGG GCT GGG GTG GCA G-3' | 74 bp |

1) Primers recognize endogenous and transgenic mRNAs. All primers ordered from Sigma.

The results were analyzed with the comparative Ct method ($\Delta\Delta$Ct), normalizing the target Ct to a housekeeping gene (Cyclophilin G) and relative to a calibrator standard, and calculated

with $2^{-\Delta\Delta Ct}$. For statistics, error bars were presented as standard error of the mean (SEM), and p values were calculated with one-way ANOVA and Dunnett's multiple comparisons test as well as unpaired t test, where $p < 0.05$ was used as the statistical significance. Statistics were calculated using the GraphPad Prism 7 software.

## 3.10. Embryoid Body Assay

To initiate embryoid body (EB) formation, cells were treated with 0.5 mM EDTA, then scraped of the culture dish with a pipet tip and transferred to ultra-low attachment dishes (Corning, 3471) in hES medium without bFGF (KnockOut DMEM (Life Technologies, 10829018), 20% KO serum replacement (Life Technologies, 10828028), 1% GlutaMAX (Life Technologies, 35050-038), 0.1 mM β-mercaptoethanol (Life Technologies, 31350-010), 1% nonessential amino acids (Life Technologies, 11140050)) supplemented overnight with 5 µM Rho-associated protein kinase (ROCK) inhibitor (Y-27632; Selleckchem, S1049) to help cell survival. Medium was changed every other day by centrifugation (200 g, 5 min, RT), and suspending EBs to new medium. New medium was added every other day. EB clusters were disrupted by pipetting or by cutting with a scalpel. EBs were grown in suspension for 2 to 4 weeks and plated to cell culture dishes coated with 0.1 % gelatin (Sigma, G1890), where they were cultivated until enough outgrowth had formed for immunocytochemistry to detect germ layer markers.

One third of the EBs were transferred to grow in fibroblast medium 5 days after initiating the EB formation to direct differentiation towards endodermal cell types, and these EBs were treated similarly to the ones in hES without bFGF.

## 3.11. PCR for episome detection

DNA samples for PCR detection of residual reprogramming and immortalization episomes were collected as follows: cells were detached from the plate by scraping with a pipet tip, transferred to a microcentrifuge tube, centrifuged at 200 g, 5 min, RT, and cell pellet was stored to -80 °C. To obtain DNA from collected cell pellets, the pellets were lysed by adding 100 µl DirectPCR Lysis Reagent

(Viagen, 302-C) with 1 µg/µl proteinase K (Macherey-Nagel, 740506), and incubating them in 55 °C, overnight, in a rotating Shake 'n' Stack Hybridization Oven (Hybaid). Lysis mixture was ready to use after inactivation by incubating in 85 °C for 1 hour. PCR reaction mixtures were prepared per one reaction (à 20 µl) as follows:

| | |
|---|---|
| 4 µl | 5X Phusion HF green buffer (Thermo Scientific, F-534) |
| 1.6 µl | 2.5 mM dNTPs (Promega, U1420) |
| 0.2 µl | 100 % DMSO (Thermo Scientific, F-534) |
| 1 µl | 10 µM forward primer |
| 1 µl | 10 µM reverse primer |
| 0.2 µl | 2U/µl Phusion DNA Pol (Thermo Scientific, F-534) |
| 4 µl | 5 M Betaine (Fluka, 14300) |
| 7.5 µl | $H_2O$ (Lonza, 51200) |
| 0.5 µl | DNA from lysis mixture |

iPSC lines were assayed for OriP, EBNA-1 and dCas9 genes, and GAPDH was used as an internal control to ensure that all samples contained genomic DNA. pCXLE-dCas9VP192-GFP-shp53 plasmid was used as a positive control for EBNA-1, OriP and dCas9 at 0.2 ng/µl concentration. DNA samples from all the lymphoblastoid cell lines were used as controls. Since the LCLs had been immortalized with EBV, they should contain EBV-related genes, i.e. OriP and EBNA-1. HFF sample was used as a negative control. Primers used in episome PCR are listed in table 5.

Table 5. Primers for episome detection PCR.

| Gene | Forward primer[1] | Reverse primer[1] | Product length |
|---|---|---|---|
| OriP | pEP4-SF1-oriP<br>5'-TTC CAC GAG GGT AGT GAA CC-3' | pEP4-SR1-oriP<br>5'-TCG GGG GTG TTA GAG ACA AC-3' | 544 bp |
| EBNA-1 | pEP4-SF2-oriP<br>5'-ATC GTC AAA GCT GCA CAC AG-3' | pEP4-SR2-oriP<br>5'-CCC AGG AGT CCC AGT AGT CA-3' | 666 bp |
| dCas9 | Cas9_F1<br>5'-AAA CAG CAG ATT CGC CTG GA-3' | Cas9_R2<br>5'-CTG TCT GCA CCT CGG TCT TT-3' | 1934 bp |
| GAPDH | GAPDH h-m F1 sg<br>5'-AAG AAG GTG GTG AAG CAG GC-3' | GAPDH h-m R1 sg<br>5'-CAG GAA ATG AGC TTG ACA AAG-3' | 164 bp |

1) Primers ordered from Sigma.

PCRs were run according to following programs:

OriP and EBNA-1

| | | | |
|---|---|---|---|
| | Initial denaturation | 98 °C | 3 min |
| 8 cycles: | Denaturation | 98 °C | 10 sec |
| | Annealing (down 0.5 °C/cycle) | 66 – 62 °C | 30 sec |
| | Elongation | 72 °C | 20 sec |
| 30 cycles: | Denaturation | 98 °C | 10 sec |
| | Annealing | 62 °C | 30 sec |
| | Elongation | 72 °C | 20 sec |
| | Final elongation | 72 °C | 8 min |

GAPDH

| | | | |
|---|---|---|---|
| | Initial denaturation | 98 °C | 3 min |
| 8 cycles: | Denaturation | 98 °C | 10 sec |
| | Annealing (down 0.5 °C/cycle) | 65 – 61 °C | 30 sec |
| | Elongation | 72 °C | 7 sec |
| 30 cycles: | Denaturation | 98 °C | 10 sec |
| | Annealing | 61 °C | 30 sec |
| | Elongation | 72 °C | 7 sec |
| | Final elongation | 72 °C | 8 min |

dCas9

| | | | |
|---|---|---|---|
| | Initial denaturation | 98 °C | 3 min |
| 8 cycles: | Denaturation | 98 °C | 10 sec |
| | Annealing (down 0.5 °C/cycle) | 68 – 64 °C | 30 sec |
| | Elongation | 72 °C | 1 min |
| 30 cycles: | Denaturation | 98 °C | 10 sec |
| | Annealing | 64 °C | 30 sec |
| | Elongation | 72 °C | 1 min |
| | Final elongation | 72 °C | 8 min |

Annealing temperatures were calculated with the New England Biolabs web-based $T_m$ Calculator (NEB 2017).

PCR products for OriP, EBNA-1 and GAPDH were run in 1.0 – 1.5 % agarose (Fisher Bioreagents, 10766834) Tris/Borate/EDTA (TBE; Life Technologies, 15581044) gel, supplemented

with Midori Green Advance (NIPPON Genetics Europe, MG04), and PCR products for dCas9 were run in 1.0 – 1.2 % agarose Tris/Acetate/EDTA (TAE; Fisher Bioreagents, B1335-1) gel, supplemented with Midori Green Advance. GeneRuler™ 1 kb Plus DNA Ladder (Thermo Scientific) was used as the molecular ladder. Gels were imaged with Bio-Rad Gel Doc™ XR+ System and Image Lab™ Software.

### 3.12. iPSC freezing

iPSC lines were frozen as follows: cells were detached from plates with similar EDTA treatment as in passaging, and cells were scraped with a cell scraper and 500 µl of Freezing medium I (DMEM (Sigma, D6546), 20 % FBS (Life Technologies, 10270-106)), and transferred to a freezing vial (Fisher Scientific, 12-567-501). 500 µl of cold Freezing medium II (DMEM (Sigma, D6546), 20 % FBS (Life Technologies, 10270-106), 20 % DMSO (Sigma, D2650)) was added drop-wise, and the cells and media were mixed by turning the vial a few times. Vials were stored to -80 °C in styrofoam freezing boxes for initial freezing, and then transferred to -150 °C for long-time storage.

# 4. Results

## 4.1. Selection of lymphoblastoid cell line for CRISPRa mediated reprogramming

The CRISPRa mediated reprogramming method established in HFFs had had a lower reprogramming efficiency than the transgenic system. For this reason, it was important to select an LCL with the highest possible reprogramming efficiency for the first CRISPRa reprogramming experiments. To determine which LCL had the highest reprogramming efficiency, all cell lines were reprogrammed with transgenic reprogramming factors. LCL IB-D5 gave rise to the highest number of iPSC colonies, quantified by AP staining (figure 1). To determine whether the transfection efficiency of the LCLs caused the variability in the reprogramming efficiencies, a transfection efficiency assay was performed. All LCLs were transfected with a GFP plasmid, and three days later the percentage of GFP positive cells was determined with FACS. The percentage was used as a direct indicator of the transfection efficiency. All cell lines had similar transfection efficiencies between 55 – 80 %, and no correlation between the reprogramming and transfection efficiency was detected, with $R^2 = 5{,}7*10^{-5}$ (figure 1). This suggested that the cell lines had different reprogramming efficiencies, and as IB-D5 had the highest efficiency, it was selected for the first reprogramming experiments.

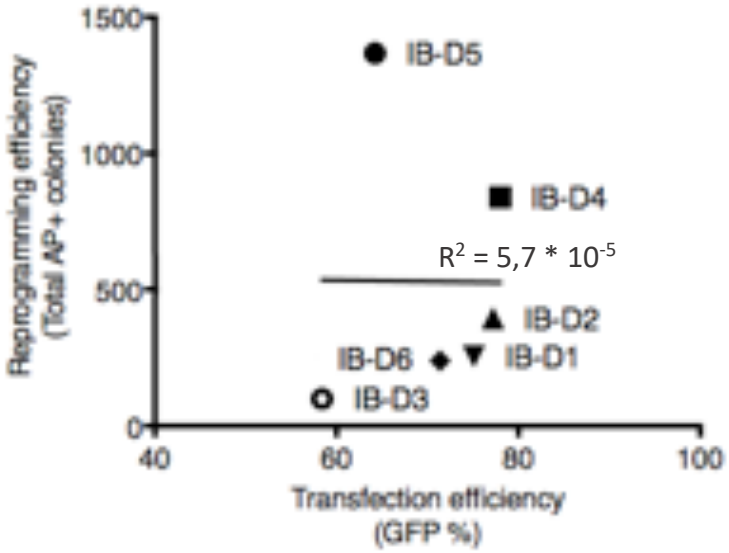



Figure 1. Transfection versus reprogramming efficiencies of the six LCLs. Linear trend line drawn in black. Data from two individual experiments, N = 1 for both experiments.

## 4.2. CRISPRa mediated reprogramming of IB-D5

Since LCLs had never been reprogrammed using the CRISPRa system, seven different gRNA combinations were tested to see if they would be able to reprogram the cells. Each CRISPRa combination included the dCas9-VP192 activator with shp53 and gRNAs targeting the promoter regions of reprogramming factors OCT4, SOX2, KLF4, C-MYC and LIN28 (OSKML), on top of which additional gRNAs for REX1, NANOG, LIN28 and the EEA-motif (either 1 or 5 guides) were assayed (table 6). Transgenic GFP was used as the negative control and a combination of transgenic OCT4, SOX2, KLF4, L-MYC and LIN28 with shp53 as the positive control.

Table 6. CRISPRa combinations and controls

| Combination | Transgenes | Guide-RNAs[1] | | | | | |
|---|---|---|---|---|---|---|---|
| | | OSKML | R | N | L | E1 | E5 |
| CRISPRa[2] | | | | | | | |
| G | GFP | x | | | | | |
| E1 | | x | | | | x | |
| E5 | | x | | | | | x |
| RN | | x | x | x | | | |
| RNE1 | | x | x | x | | x | |
| RNE5 | | x | x | x | | | x |
| NLE5 | | x | | x | x | | x |
| Controls | | | | | | | |
| Negative | GFP | | | | | | |
| Transgenic | OCT4, SOX2, KLF4, L-MYC, LIN28, shp53 | | | | | | |

1) Guide-RNAs targeting the promoter regions of the endogenous genes
2) All CRISPRa combinations included the dCas9-VP192 activator and shp53
OSKML = OCT4, SOX2, KLF4, C-MYC, LIN28, R = REX1, N = NANOG, L = LIN28, E1 = one guide for EEA motif, E5 = five guides for EEA motif

LCL IB-D5 was transfected with the seven CRISPRa combinations and the controls, and the cells were grown as described in the figure 2. The emerged iPSC colonies were recognized based on morphology and the number of iPSC colonies was quantified with AP staining (figure 2).

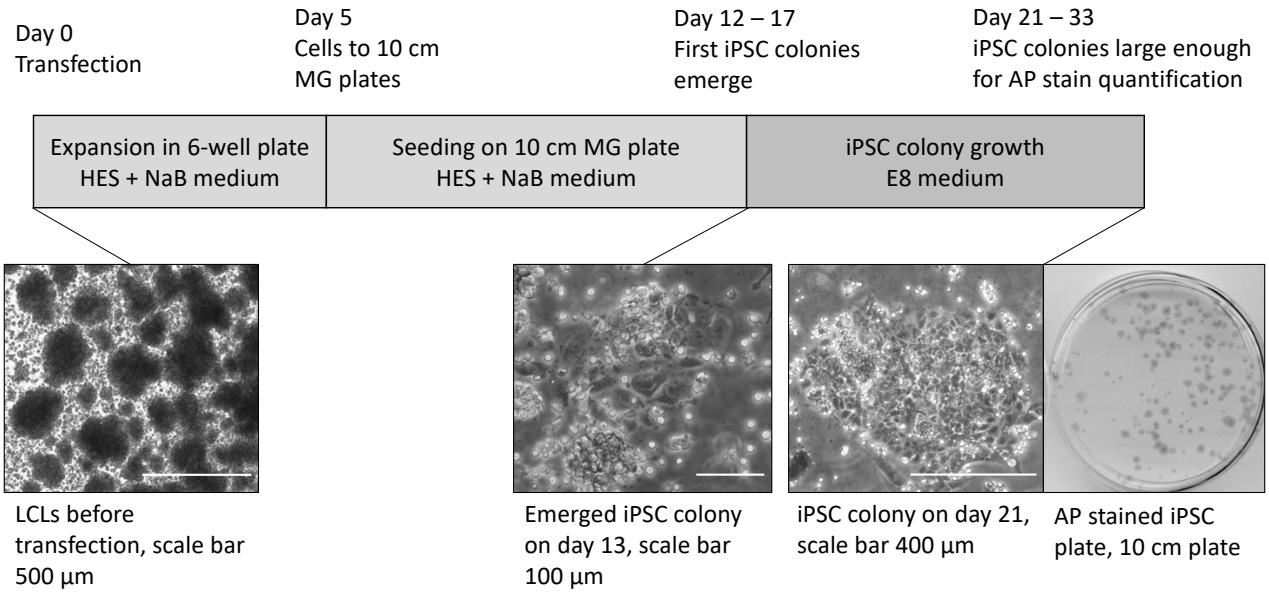



Figure 2. Reprogramming of lymphoblastoid cell lines. iPSC = induced pluripotent stem cell, AP = alkaline phosphatase, MG = matrigel, LCLs = lymphoblastoid cell lines

Three CRISPRa combinations, which all included the targeting of the EEA-motif, gave rise to iPSC colonies (figure 3). Combination of guides for OSKML with REX1, NANOG and one guide for the EEA motif (RNE1) gave rise to the highest number of iPSC colonies. However, the efficiency was low with the mean efficiency of 0.00018 % from three inductions for combination RNE1 compared to the transgenic 0.00543 % from two inductions.

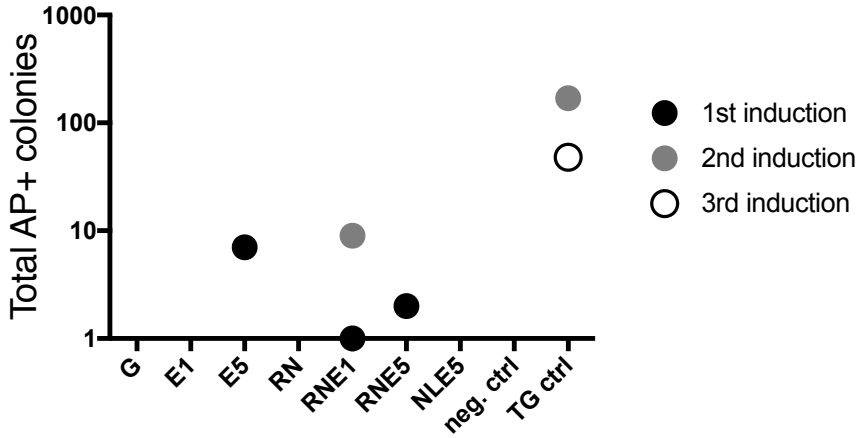



Figure 3. Number of AP positive colonies yielded from reprogramming of IB-D5 with reprogramming plasmid combinations. Data from three individual experiments, combination NLE5, the negative control and the transgenic (TG) control were not included in the first experiment.

## 4.3. Guide-RNA function validation

Because of the low reprogramming efficiency, a gRNA function validation assay was performed to determine whether the used gRNAs were able to induce the gene expression of the targeted genes in LCLs. IB-D5 was transfected with the CRISPRa combinations, and after four days, the expression levels of the targeted genes were analyzed with RT-qPCR. The expression levels elicited by the CRISPRa combinations showed a consistent trend of activation compared to the negative control for all targeted genes, except for LIN28 (figure 4). However, there were only a few statistically significant gene activation levels, determined by One-way ANOVA and Dunnett's multiple comparisons test between the CRISPRa combinations and the negative control.

The gRNA function validation assay also showed whether the gene expression levels of OCT4, SOX2, KLF4 and LIN28 elicited by the CRISPRa combinations were similar to those in the transgenic control. OCT4 was expressed similarly with both CRISPRa and transgenic systems, and no statistical significance was found between the CRISPRa combinations and the transgenic control, determined by One-way ANOVA and Dunnett's multiple comparisons test. This suggested that the CRISPRa system can elicit as high gene expression levels as the transgenic system. However, the expression levels for SOX2, KLF4 and LIN28 were much higher in the transgenic control ($P < 0.0005$, $P < 0.0001$, $P < 0.0001$ respectively for all comparisons) potentially explaining the generally low reprogramming efficiency of the CRISPRa system.

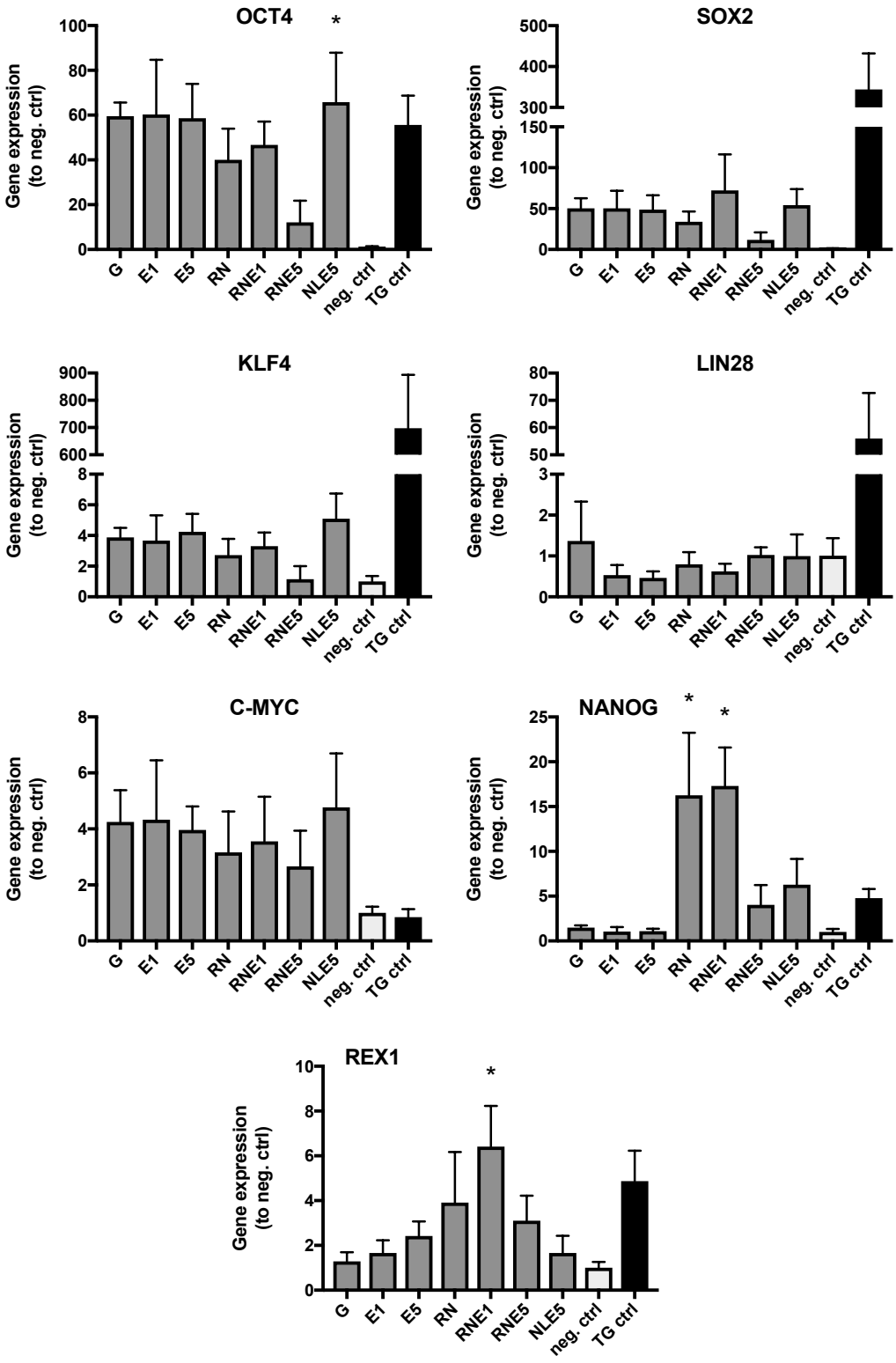



Figure 4. Relative expression of the CRISPRa targeted genes in IB-D5 on day 4 after transfection with the reprogramming combinations. Transgenic control does not include C-MYC, NANOG or REX1. Expression levels presented as a fold change, normalized to negative control. Error bars represent SEM. Statistical significance between the CRISPRa combinations and the negative control determined with ordinary one-way ANOVA and Dunnett's multiple comparisons test. * $P < 0.05$. N = 3.

Since this was the first time the CRISPRa reprogramming was established in LCLs, there were no LCL models which could have been used as the positive control to show the necessary gene activation by CRISPRa for efficient reprogramming. For this reason, the CRISPRa elicited gene expression levels in IB-D5 were compared to those in HFFs, in which the CRISPRa reprogramming had already been successfully established with the same gRNAs (Weltner et al. 2018). The gRNA function validation assay was performed in HFFs alongside IB-D5 with the negative and positive controls and the combination RNE1. RNE1 was selected because it gave rise to the highest number of iPSC colonies in IB-D5 in the reprogramming experiment and elicited the highest expression of NANOG and REX1.

None of the genes were significantly upregulated or downregulated by the CRISPRa system in HFFs compared to the negative control, determined with a two-way t test (figure 5). Since the assay could only be performed twice in HFFs, the statistical significance was affected. These results, nevertheless, showed similar activation trends for OCT4, SOX2, LIN28, NANOG and REX1, while it seems that in HFFs the CRISPRa system rather repressed the expression of KLF4 than activated it as in IB-D5. However, KLF4 had high expression already in control HFFs, which might affect the activation levels. C-MYC seemed to be unaffected by the CRISPRa combination. None of the genes, which were targeted by CRISPRa and included in the transgenic control, were activated with the CRISPRa system to the same level as they were expressed in the transgenic control. However, no statistical significance between the expression levels of the genes activated by the CRISPRa system and expressed by the transgenic control was found, most likely due to the number of replications.

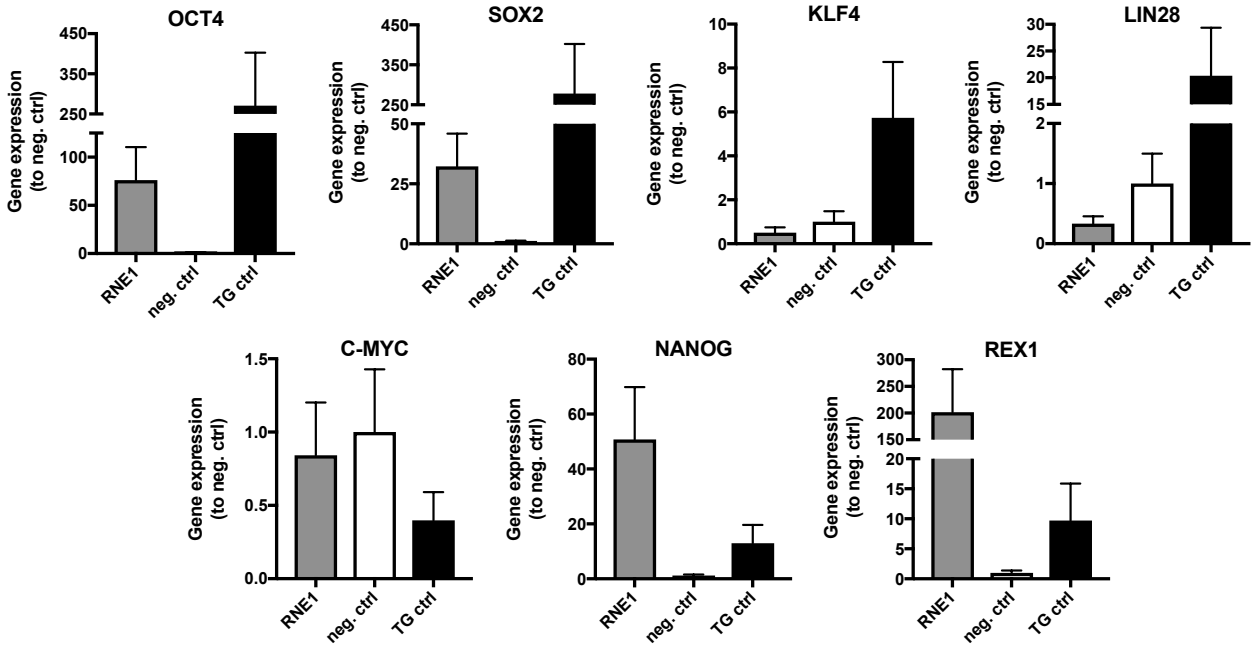



Figure 5. Relative gene expression of the CRISPRa targeted genes in HFFs four days after transfection with combination RNE1 and the controls. Expression levels presented as a fold change, normalized to the negative control. Error bars represent SEM. No statistical significance was found between the expression levels between RNE1 and negative control or between RNE1 and transgenic control, determined with a two-way t test. N = 2.

To compare the expression levels elicited by the CRISPRa system in IB-D5 and HFFs, the transgenic expression levels of OCT4 were first compared to determine whether the two cell types had been transfected with similar efficiencies. OCT4 was expressed at similar levels endogenously in the negative controls in both cell types, indicating that differences in the expression levels of OCT4 in the transgenic control would result from different transfection efficiencies between the cell lines. The OCT4 expression was similar in both cell types (figure 6), which suggested that the cell types had been transfected similarly, and that the gene expression levels elicited by the CRISPRa combination RNE1 could be compared.

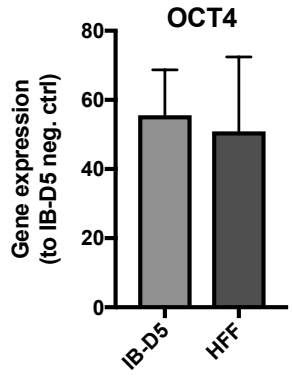



Figure 6. Relative expression of OCT4 in the transgenic control in IB-D5 and HFFs four days after transfection. Expression levels presented as a fold change, normalized to the negative control in IB-D5. Error bars represent SEM. P = 0.8540, determined by a two-way t test. N = 3 for IB-D5, N = 2 for HFF.

There was no statistical significance between the expression levels of OCT4, SOX2, LIN28, C-MYC or NANOG elicited by the CRISPRa combination in IB-D5 and HFFs, determined with a two-way t test (figure 7). However, the gene expression levels of KLF4 and REX1 were significantly higher in HFFs than in IB-D5.

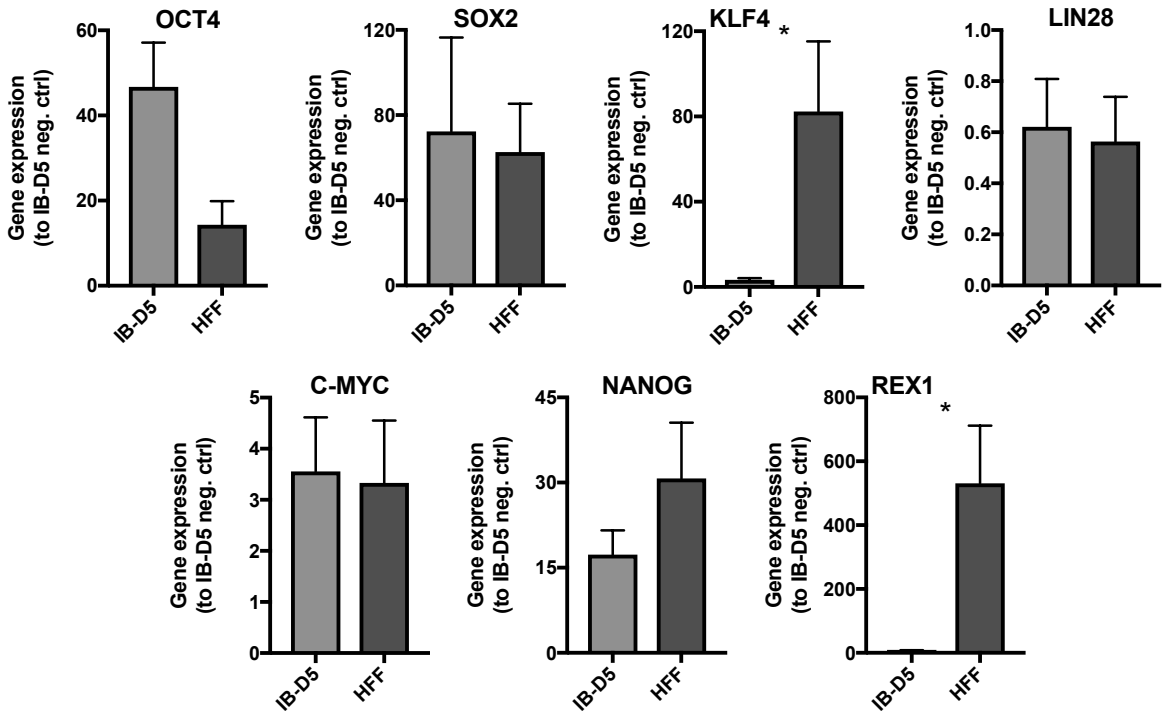



Figure 7. Relative gene expression levels of the CRISPRa targeted genes in IB-D5 and HFFs four days after transfection with the combination RNE1. Expression levels expressed as fold change, normalized to the negative control in IB-D5. Error bars represent SEM. Statistical significance determined with a two-way t test. * $P < 0.05$. N = 3 for IB-D5, N = 2 for HFFs.

Overall, these results suggest that the gRNAs were suboptimal for efficient expression of some of the targeted genes, such as KLF4 and LIN28, in IB-D5. It also seems that some of the gRNAs, for example the ones for REX1, function better in HFFs.

## 4.3. Derivation and characterization of iPSC lines from five LCLs

Based on the results from the CRISPRa reprogramming experiment and the gRNA function validation assay, combination RNE1 was selected as the best for reprogramming of LCL IB-D5. To see whether it would also reprogram other LCLs, all six LCLs were transfected with the combination. After first induction, only three LCLs generated iPSC colonies (figure 8). The induction was repeated for all lines, and on the second time five LCLs gave rise to iPSC colonies. IB-D3 was included only in the second induction set due to low proliferation rate. There was a high variability in reprogramming efficiencies, and for example IB-D2 did not yield any iPSC colonies in the first induction, while in the second induction it had the highest reprogramming efficiency, giving rise to over 650 iPSC colonies.

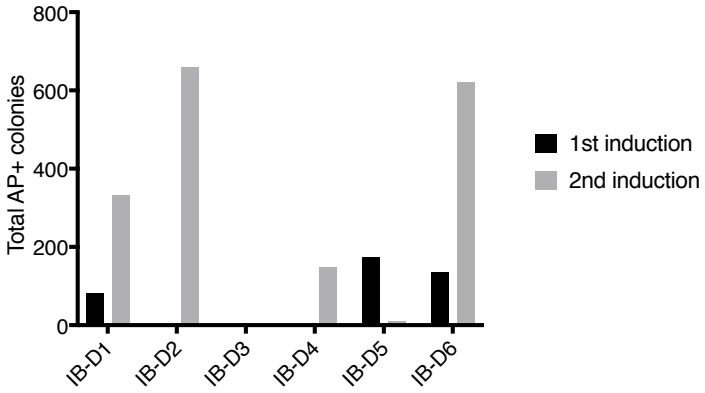



Figure 8. Number of AP positive colonies from two induction sets of the six LCLs, reprogrammed with CRISPRa combination RNE1. IB-D3 was included only in the second induction set.

To prove that the CRISPRa system was able to reprogram LCLs into bona fide iPSCs, three iPSC line clones of each LCL, except for IB-D3, were derived, named and characterized according to table 7. Lines were characterized with basic characterization standards, and one clone of each established iPSC line was characterized with all the assays.

| Table 7. Characterization of LCL derived iPSC lines. | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| iPSC line | LCL | Clone | PP ICC[1] | PP marker ICC | | | PP RT-qPCR[1] | EB assay[1] | EB germ layer marker ICC | | | Episome gene PCR[1] | ID PCR[1] |
| | | | | Oct4 + Tra-1-60 | Lin28 + Tra-1-81 | Sox2 + Tra-1-81 | | | α-SMA | TUBB3 | SOX17 | | |
| HEL159 | IB-D6 | 3 | 3 | x | x | | 9 | 3 | x | x | x | 5 | 4 |
| | | 6 | 3 | x | x | | 8 | 3 | x | x | x | 5 | 4 |
| | | 9 | 3 | x | x | | 8 | 3 | x | x | x | 5 | 4 |
| HEL160 | IB-D2 | 4 | 3 | x | x | | 10 | 6 | x | x | x | 4 | 7 |
| | | 12 | 4 | x | x | | | 5 | x | x | | 6 | 3 |
| | | 16 | 3 | x | x | | | | | | | 6 | 5 |
| HEL161 | IB-D1 | 4 | 4 | x | x | | 8 | 5 | x | x | x | 5 | 6 |
| | | 15 | 4 | x | x | | | 5 | x | | | 3 | 3 |
| | | 17 | 4 | x | x | | | 5 | x | x | x | 5 | 6 |
| HEL162 | IB-D4 | 4 | 8 | x | | x | 7 | 6 | x | x | x | 3 | 3 |
| | | 8 | 5 | x | | x | 6 | | | | | 3 | 7 |
| | | 9 | 5 | x | | x | | 6 | x | x | x | 3 | 3 |
| HEL163 | IB-D5 | 8 | 4 | x | | x | 6 | 5 | x | x | x | 5 | 4 |
| | | 9 | 4 | x | | x | 7 | | | | | 5 | 4 |
| | | 11 | 4 | x | | x | 7 | 6 | x | x | | 5 | 4 |

1) Passage number when a sample was collected or an assay began

iPSC = induced pluripotent stem cell, LCL = lymphoblastoid cell line, PP = pluripotency, ICC = immunocytochemistry, RT-qPCR = reverse transcriptase quantitative polymerase chain reaction, EB = embryoid body, ID PCR = identification genotyping with polymerase chain reaction

All iPSC clones resembled ESCs morphologically: the cells were small, they formed tightly packed, round colonies with distinct borders, and had a high nucleus/cytosol ratio. All clones were stained with ICC for pluripotency stem cell markers OCT4, TRA-1-60, TRA-1-81, and LIN28 or SOX2 (figures 9 and S1).

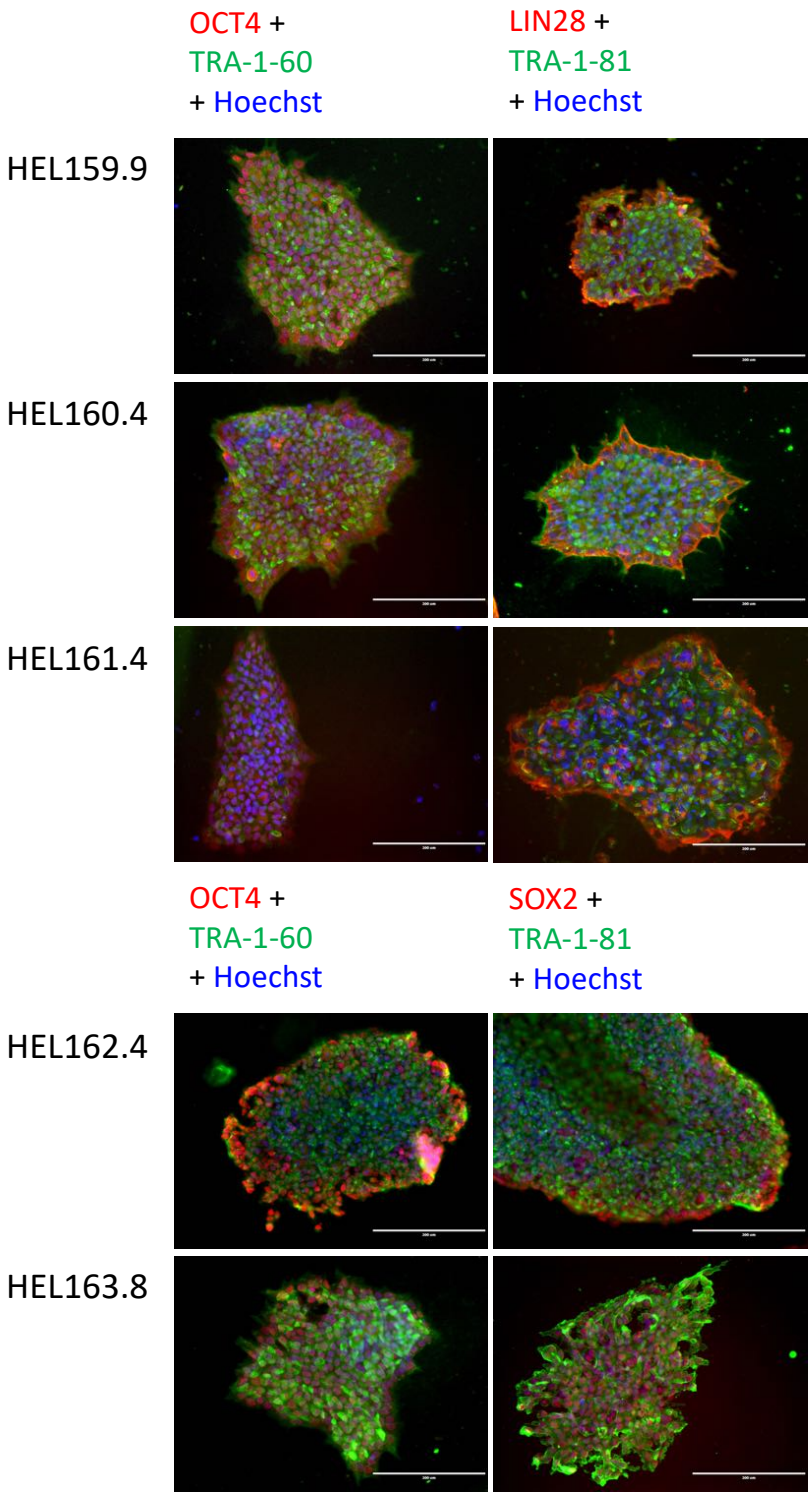



Figure 9. Pluripotency marker expression of 5 iPSC line clones assayed with ICC for OCT4, TRA-1-60, TRA-1-81, and LIN28 or SOX2. Nuclei stained with Hoechst33342. Scale bar = 200 µm.

All assayed clones were positive for the expression of pluripotency marker genes OCT4, SOX2, NANOG, REX1 and TDGF1, analyzed also with RT-qPCR (figure 10). OCT4 and NANOG were expressed consistently at a lower level in LCL-derived iPSCs compared to the HEL46.11 iPSC control. SOX2 expression, on the other hand, was slightly higher in all the LCL-derived iPSCs. REX1 and TDGF1 expression levels were similar to the ones in HEL46.11.

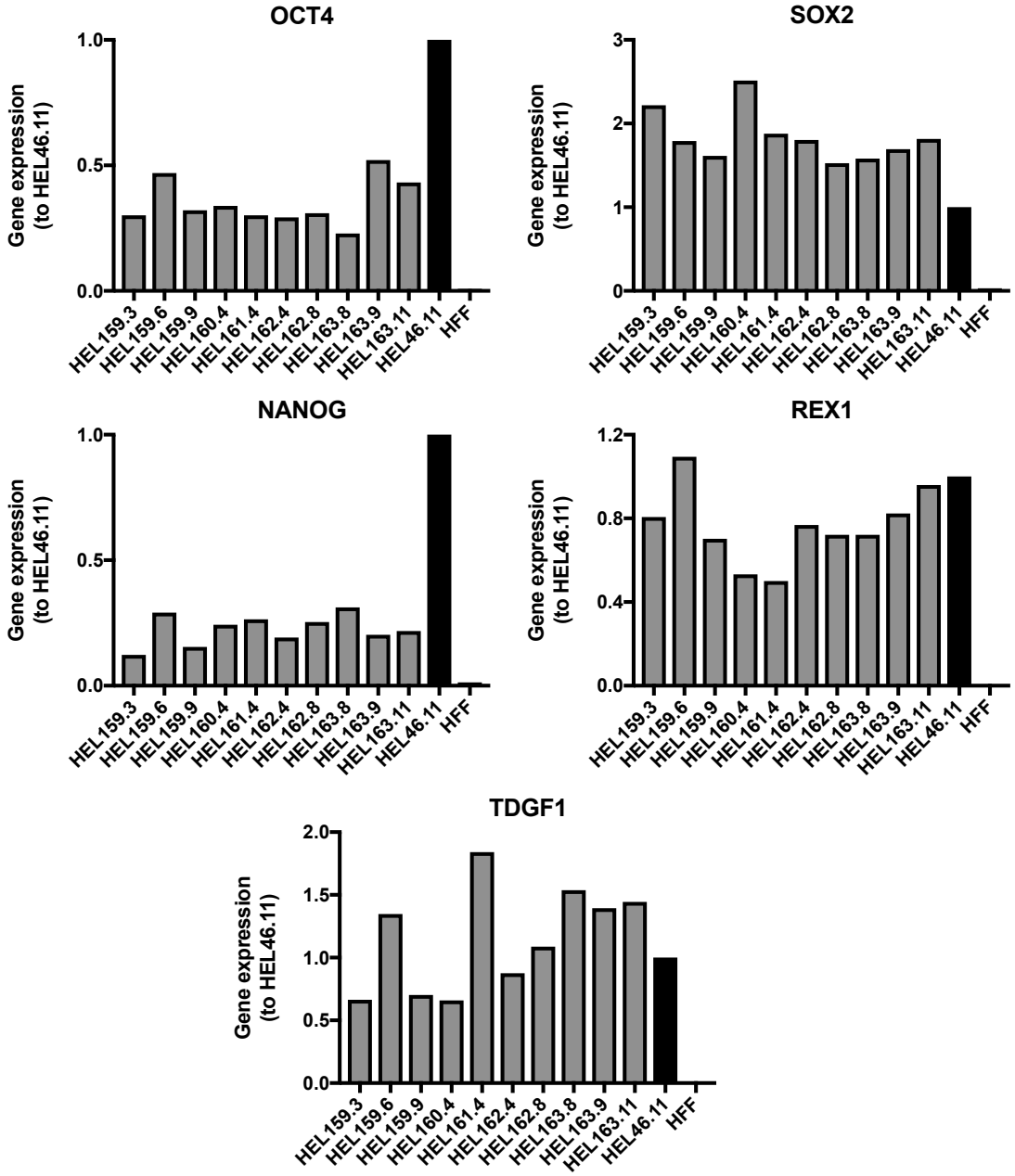



Figure 10. Relative expression of pluripotency markers OCT4, SOX2, NANOG, REX1 and TDGF1 in 10 LCL-derived iPSC lines analyzed with RT-qPCR. Expression levels are normalized to the HEL46.11 (p38) iPSC control, HFF used as a negative sample. N = 1 of each line.

The iPSC clones were analyzed for any residual reprogramming plasmids with PCR for dCas9, OriP, EBNA-1 and GAPDH genes (figure 11). 11 out of 15 iPSC clones were negative for dCas9. The absence of dCas9 suggests that the endogenous pluripotency network had stabilized in the cells, since no dCas9 mediated reprogramming gene activation could occur. Out of the four positive lines, the faint band in HEL162.4 suggests that only a portion of the cells in the sample contained the activator plasmid. HEL160.4, HEL162.4 and HEL163.11, however, gave quite intense bands for dCas9, meaning that the cells still contained transgenic plasmid sequences.

13 out of 15 iPSC clones and all five LCLs were positive for EBNA-1 and OriP. Since the LCLs had been immortalized with EBV, it was expected that they would give bands, and it might be that the EBNA-1 and OriP positive iPSC clones were still harboring the immortalization plasmids. The bands for EBNA-1 and OriP from iPSC clones might have also come from the gRNA reprogramming plasmids, and the bands in the dCas9 positive clones were at least partly due to the activator plasmid. In each cell line, the strongest band came from the LCL sample, indicating that the immortalization and reprogramming plasmids have been at least partially diluted out of the derived iPSC clones. HEL160.12 and HEL161.17 seem to have lost all episomes.

GAPDH showed similar levels of genomic DNA in all cellular samples.

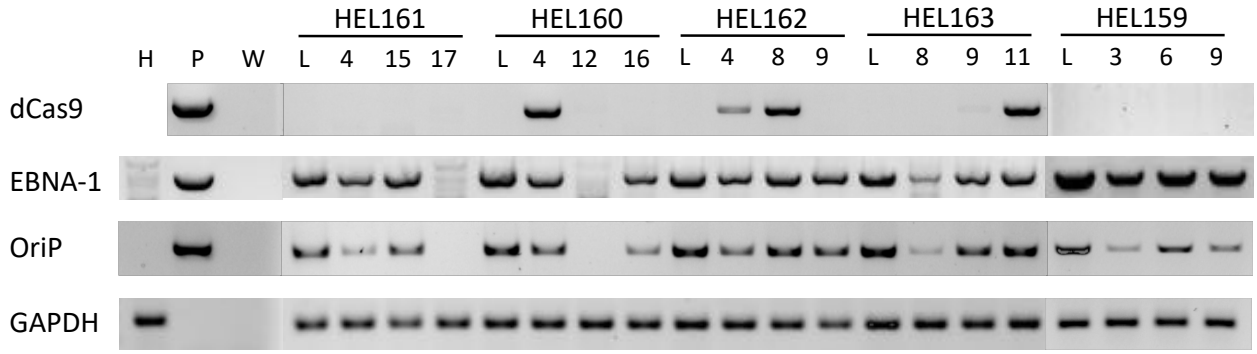



Figure 11. PCR for dCas9, EBNA-1, OriP and GAPDH of LCL-derived iPSC lines. L = parental LCL, H = HFF control, P = activator plasmid control, W = water control.

To see whether the generated iPSC lines were able to differentiate to all three germ cell layers, an embryoid body assay was performed to 12 out of 15 clones. A clear signal of tri-lineage differentiation was detected from 9 out of 12 clones with ICC (figures 12 and S2).

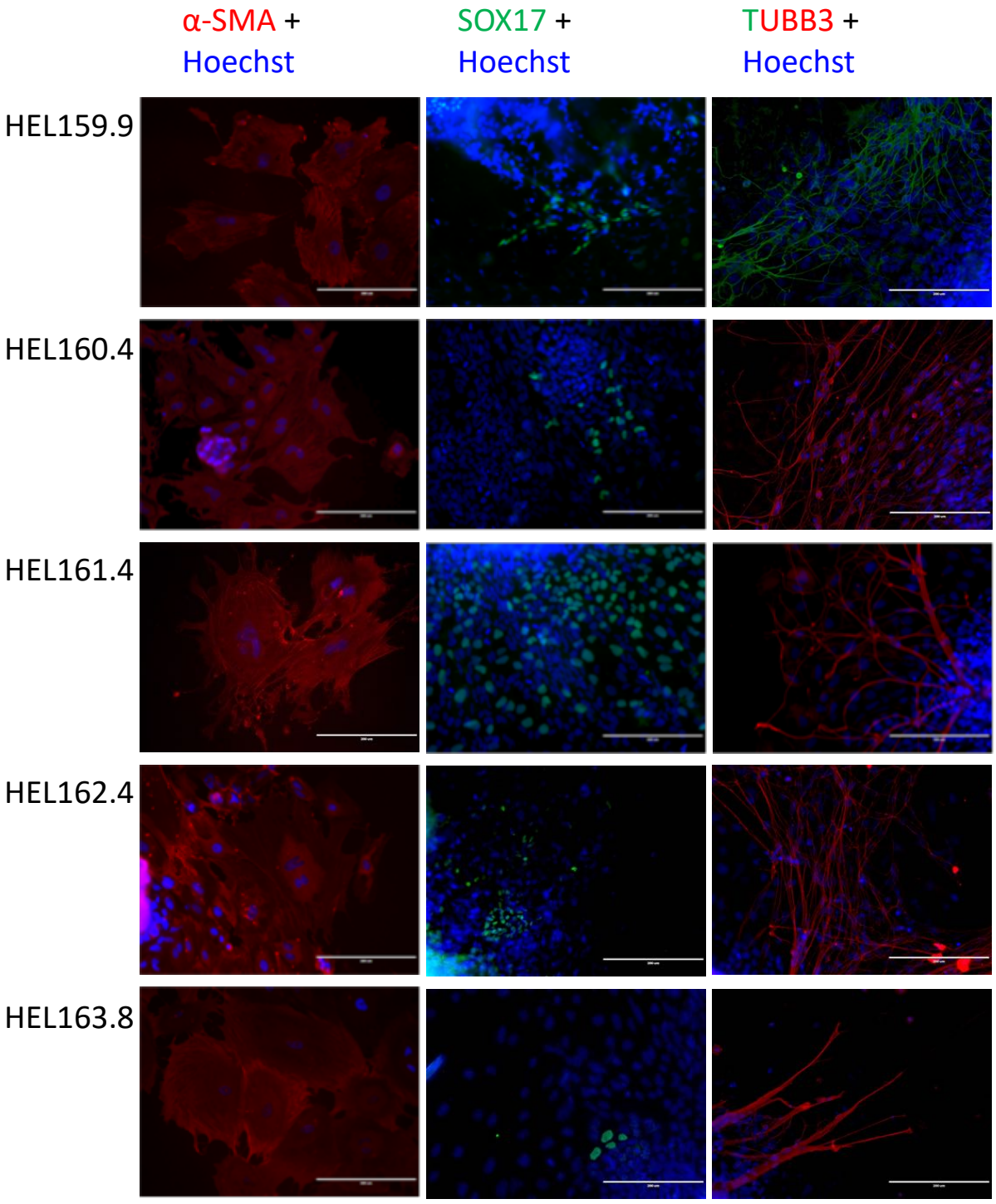



Figure 12. ICC shows clear signal of tri-lineage differentiation of 5 iPSC clones derived from different LCLs. Smooth muscle actin (α-SMA) was used as a mesendodermal marker, SOX17 as a endodermal marker, and β-tubulin (TUBB3) as an ectodermal marker. Nuclei stained with Hoechst33342. Scale bar = 200 μm.

All lines were genotyped for the D1S80 locus to confirm iPSC line identity. Each iPSC clone matched its parental LCL (figure 13). HEL161 and HEL160 had the same allele of the locus, and additional loci should be genotyped in order to verify that no mixing of the clones had occurred. Also, the genotyping showed that HEL163 (LCL IB-D5) might harbor a chromosomal abnormality, since it showed three bands of the allele instead of two, which is the maximum number of bands from cells with normal karyotypes. Since the three bands are also present in the LCL sample, the possible mutation had not emerged during the reprogramming process. Karyotyping should be performed in order to identify the chromosome anomaly.

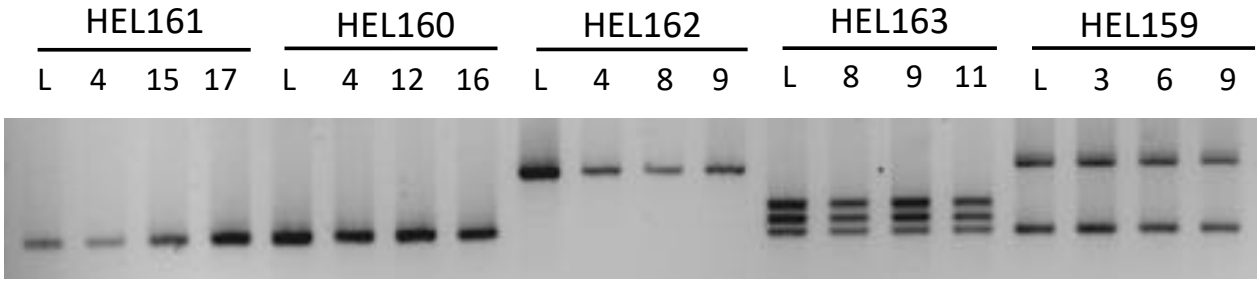



Figure 13. Genotyping of iPSC clones and LCLs for the D1S80 locus. L = parental LCL.

In conclusion, the CRISPRa mediated reprogramming was successfully established in five different LCLs and three bona fide iPSC clones of each LCL were derived. Two vials of each clone were frozen and returned to the THL biobank as a future research resource.

# 5. Discussion

In order to utilize the vast LCL collections in biobanks and cell repositories more thoroughly, it is crucial to understand the cell specific events that take place in LCLs during the reprogramming process, and to this end the CRISPR/dCas platform has been applied. In this thesis project, the novel CRISPRa mediated reprogramming method was established for the first time in biobanked LCLs. The success of the established method was demonstrated by the reprogramming of five LCLs into bona fide iPSC lines. The reprogramming efficiencies were varying between experiments, and we showed that the gRNAs used in these experiments were suboptimal for efficient gene activation of some of the targeted genes in LCLs. These results also raised a question of the utility of LCLs in iPSC technology derived research, given the possible chromosomal aberration detected in one of the cell lines. This thesis work serves as the first stepping stone on the way to understand the events occurring in LCLs during reprogramming and to develop an LCL specific reprogramming method, which would result in better quality iPSCs and higher reprogramming efficiencies.

## 5.1. CRISPRa reprogramming of LCLs

The CRISPRa mediated reprogramming method was successfully employed to reprogram five out of six biobanked LCLs, demonstrated by the characterization of the derived iPSC lines. All clones met the standard characterization requirements but given the time and resource constraints, complete characterization could not be performed. The samples were collected at early passages, which might have impacted the detection of the residual episomal plasmids and the pluripotency marker expression levels. It is likely that that the reprogramming and immortalization plasmids would have been diluted out with further passaging, as previous studies have shown that LCL-derived iPSCs are free of EBNA-1 and OriP sequences by passage 20 (Barrett et al. 2014; Rajesh et al. 2011; Thomas et al. 2015). The cell lines which were positive for the episomal plasmid sequences should be passaged further and re-analyzed. Also, samples from later passages might have had more similar pluripotency gene expression levels as the HEL46.11 control from passage 38, since it has been shown that early and late passage iPSCs exhibit differential gene expression (Chin et al. 2012), and early passage iPSCs can still exhibit cell-type-of-origin-specific epigenetic modifications (Polo et al.

2010). In the light of this, it would have been interesting to use an LCL-derived, transgenically reprogrammed, earlier passage iPSC line as a control for the characterization experiments, but due to resource constraints such cell lines could not be derived. To fully characterize the generated iPSC lines, the following assays should be performed (Martí et al. 2013): the lines should be karyotyped to confirm the genomic integrity; the lines should be genotyped with additional markers; and a pluritest should be performed to verify the pluripotent state of the cell lines in more detail.

Even though the CRISPRa mediated reprogramming method was proven to successfully reprogram LCLs into bona fide iPSCs, the reprogramming efficiency was generally low and variable between experiments. The variation could be caused by several factors, such as the cell proliferation of the LCLs during electroporation, the success of the electroporation, handling of the cells during media change, and differences in the time period the cells were grown in hES and E8. The efficiency of transgenic LCL reprogramming has been improved by the optimization of the transfection and culturing protocols (Kumar et al. 2016), and a more optimized protocol could also improve the CRISPRa reprogramming efficiency and diminish the variation. However, since the generated iPSCs provide a perpetual source of self-renewing cells, the reprogramming method might not need to have the highest possible efficiency, if it produces high-quality iPSCs which can be expanded.

The functional assay for the CRISPRa system, i.e. the actual reprogramming of LCLs into iPSCs, proved that the CRISPRa system is capable of activating the necessary genes for successful reprogramming. However, the gRNA function validation assay showed that the gRNAs did not elicit as efficient gene activation of all targeted genes in LCLs as they did in fibroblasts, which might explain why the reprogramming efficiencies were low in LCLs. One of the factors that might more notably affect the reprogramming efficiency is the poor activation of KLF4, which has been shown to be important in B lymphocyte reprogramming (Wen et al. 2016; Di Stefano et al. 2016). Also, the LIN28 expression levels seemed to be unaffected by the CRISPRa system, but similar results have been obtained in previous studies (Weltner et al. 2018), where they showed with ICC that LIN28 is indeed expressed in some cells, but not detected in the RT-qPCR results due to the bulk effect of the sample. Similar staining assay should be performed to determine whether the low gene activation levels in these assays were also due to bulk effect. Most of the gene expression levels elicited by the CRISPRa system fell far from those of the transgenic system, except for OCT4. The high activation level of OCT4 demonstrated that it is possible to activate the endogenous gene

expression with the CRISPRa system to the same level as it is expressed in the transgenic system. However, it is not known whether the CRISPRa mediated reprogramming process requires the endogenous reprogramming factors to be expressed at the same level as they are expressed from the ectopic plasmids in transgenic reprogramming.

Comparison between the expression levels of the targeted genes in LCLs and fibroblasts revealed that the CRISPRa system had different responses in different cell types, for example the elicited REX1 expression was higher in fibroblasts than in LCLs. This could be explained by different chromatin architecture in different cell types, which can make some target loci less accessible in LCLs. Interestingly, the targeting of KLF4 in fibroblasts seemed to downregulate its expression rather than upregulate it, though statistically the downregulation was not significant. Still, the expression level of KLF4 was some 20 times higher in fibroblasts compared to LCLs. This suggests that for fibroblasts, the activation of KLF4 is not as crucial as it might be for LCLs since the base level expression is already high enough. This goes along with findings showing that KLF4 poises B lymphocytes for efficient reprogramming (Di Stefano et al. 2016). Since the gRNA function validation could only be performed twice in fibroblasts, the trends detected from this experiment should be confirmed with additional replicates. It would have been interesting to fully reprogram fibroblasts with the CRISPRa system to see if the results from the gRNA function validation would be reciprocated in the number of emerged iPSC colonies.

Based on the reprogramming experiments, the EEA motif targeting is crucial for the successful reprogramming of LCLs with the CRISPRa system. The importance of the EEA motif targeting was further supported by the gRNA function validation assay. Combinations G, E1 and E5 elicited very similar gene expressions of all the targeted genes, yet the only one that gave rise to iPSCs in the functional assay was combination E5, which included five gRNAs for the EEA motif. Also, combination RN activated all targeted genes better than RNE5, which elicited the lowest expression of nearly all of the genes, but still out of the two, only RNE5 gave rise to iPSCs, highlighting the necessity of the EEA motif targeting. The EEA motif targeting has been shown to improve the activation of NANOG and REX1 when they are targeted by the CRISPRa in fibroblasts (Weltner et al. 2018). The levels of NANOG and REX1 from combination RN and RNE1 support this, but RNE5, which theoretically should have the highest NANOG and REX1 levels, does not. This could be explained by too many EEA motif gRNA-dCas9 complexes interfering with the binding of the NANOG and REX1 specific gRNA-dCas9 complexes. The expression levels of the other targeted genes were not affected

by the EEA motif targeting, suggesting that the importance of the motif targeting in LCLs does not result from facilitating the activation of these genes. It is possible that since LCLs are transformed from fully mature B lymphocytes, their silenced pluripotency associated loci require the additional opening, induced by the EEA motif targeting throughout the gene network related to EGA, for the endogenous reprogramming factors to be able to bind after they have been activated by the gRNA-dCas9 complexes. Since the efficiencies in the first reprogramming experiment were quite low, it would be important to perform both the reprogramming as well as the gRNA function validation assays again simultaneously, in order to obtain more comparable results.

## 5.2. The usability of LCLs in iPSC technology

Previous records of LCL reprogramming have resulted in bona fide iPSCs with normal karyotypes and complete loss of the immortalization and reprogramming plasmids (Barrett et al. 2014; Rajesh et al. 2011; Thomas et al. 2015; Kumar et al. 2016; Son et al. 2017), and these kind of LCLs have been used for disease modeling (Fujimori et al. 2016). However, the results from the characterization assays raise the question whether all LCLs are safe to use in iPSC technology related research. The LCLs used in this study were obtained from THL biobank without information about their passage number or immortalization status. The genotyping assay revealed a potential chromosomal aberration in LCL IB-D5, which suggests that the cell line could have undergone immortalization, since immortalization has been previously shown to be usually accompanied by aneuploidy and other chromosomal mutations (Okubo et al. 2001). It is also possible that the other LCLs carry chromosomal mutations. Karyotyping would reveal the chromosomal state of the cell lines, and in the case of abnormal karyotypes, their effects on downstream applications should be considered. Also, if these LCLs are post-immortal, there is a chance that they are tumorigenic (Takahashi et al. 2003). For the future development of the LCL reprogramming method, it should be considered whether it would be more beneficial to focus primarily on reprogramming pre-immortal LCLs, since they produce genetically integral iPSCs with wider downstream application possibilities. Pre-immortal LCLs could also provide a more homogeneous starting material since chromosomal mutations could introduce changes to the reprogramming properties of post-immortal LCLs. However, biobanks might have not addressed the immortalization status of their LCLs, and

collections might include both pre- and post-immortal cell lines or only post-immortal ones, establishing the requirement for a reprogramming method for both types of LCLs.

The PCR for residual plasmids showed that 13 iPSC clones still had EBNA-1 and OriP sequences. Even though the episomal plasmids will most likely be diluted out of the cells with further passaging, there is a possibility that the EBV genome has been integrated into the host genome. If the episomal plasmids are not diluted out after 20 passages, the possible integration could be assayed with Southern blotting. There is no record of LCL-derived iPSCs with integrated EBV genome, so the potential effects of the integration to the iPSCs are not known. It has been shown that EBNA-1 is unable to activate the gene expression of integrated EBV related genes or host genes(Kang et al. 2001), which would suggest that the genome would not affect the gene expression of the host cell. However, in B lymphocytes, the integration of the viral genome in specific locus has been suggested to contribute to lymphomagenesis (Takakuwa et al. 2004).

LCLs have an advantage over fibroblasts as a reprogramming starting material, since they are cultured in suspension. As cells start to reprogram, they attach to the plate, making it easy to distinguish between the reprogrammable and non-reprogrammable cells in the culture. The non-reprogrammable cells can be easily discarded during medium change. This offers a great possibility for more high-through-put reprogramming of the biobanked LCL collections, as the reprogrammable cells can be selected early in the process and culture without the disturbing effects of non-reprogrammable cells, possible resulting in higher yields.

## 5.3. Future perspectives

The fifteen iPSC line clones generated in this thesis project comprise the first ever iPSC line collection derived from THL biobank cells with the CRISPRa mediated reprogramming method and serve as the proof of concept that biobanked LCLs can be reprogrammed with the method, thus paving the way for the more thorough utilization of the vast LCL collections. The presented data serves as the starting point for the understanding of the reprogramming process and the development of more efficient reprogramming methods for LCLs. To improve the system, the next step would be to screen for additional factors that could improve the reprogramming efficiency, first transgenically and then design gRNAs for the best candidates. These additional factors could be C/EBP$\alpha$, which is a B

lymphocyte specific reprogramming factor (Di Stefano et al. 2014), and factors which have been associated with the EEA motif (Weltner et al. 2018). Also, additional gRNAs for the genes already tested here could be designed and screened to see if the activation levels could be improved by targeting different sites in the promoter or even in the enhancer regions. To distinguish the best gRNAs for LCLs, a CRISPR based high-through-put screening system could be employed (Chen et al. 2015; Doench 2018).

To understand the endogenous events required for reprogramming of LCLs, the effects elicited by the endogenous reprogramming factors should be studied on a single cell level. Gene expression changes occurring during transgenic reprogramming of LCLs have been studied with whole-genome RNA sequencing (Kumar et al. 2016), but since the induced cell population is heterogeneous, the events occurring in reprogrammable cells can be masked under the bulk effect. Also, in transgenic reprogramming the crucial endogenous events might be obscured under the stochastic, but reprogramming-wise unnecessary, events elicited by the transgenic factors. With single-cell RNA sequencing and the CRISPRa mediated reprogramming, the cell-type-specific reprogramming events could be studied in a single-cell level without any transgene influence.

Another approach would be to study the chromatin accessibility and its changes during the reprogramming process in LCLs with Assay for Transposase Accessible Chromatin and high-throughput sequencing (ATAC-seq). This could reveal the epigenetic changes required for the acquisition of pluripotent state, and chromatin remodeling complexes could be employed with the CRISPRa platform to assist in those changes. For example, the dCas9-SunTag-p300 system, which was used to remodel the endogenous *Sox2* and *Oct4* promoters to achieve reprogramming in mouse fibroblasts (Liu et al. 2018), could be applied also in human cells. The changes in the chromatin could also reveal more about the role of the EEA motif targeting, as its function has been hypothesized to facilitate chromatin accessibility. If the EEA motif has a similar time-dependent function as the LINE-1 elements in regulating chromatin accessibility (Jachowicz et al. 2017), it would be interesting to study the effects of transient EEA motif targeting. It could be that that the EEA motif targeting is only required in the beginning of the reprogramming process to facilitate the initial chromatin opening and binding of the other gRNA-dCas9.

The CRISPRa mediated reprogramming method relies only on the activation of the endogenous reprogramming factor genes. However, there are several factors which impede the reprogramming process and could be repressed to improve the reprogramming efficiency (Ebrahimi

2015). To further develop the CRISPRa reprogramming method, the multiplexing capabilities of the CRISPR system could be utilized also to repress factors. The CRISPR platform offers the possibility to introduce the effector domains also fused to the gRNA (Mali, Esvelt, et al. 2013), which enables both repressing and activating gRNAs targeting different sets of genes to be used in the same cell. Another way to introduce both effects into a single cell is to utilize two dCas9 proteins from different species with different effector domains fused to them. As different origin Cas9s have different PAM site specificity (Kleinstiver et al. 2015), they could be used to target different set of genes.

# 6. Acknowledgements

I want to thank my supervisor Timo Otonkoski for taking me into his research group and for the opportunity to work on such an interesting thesis topic. I would like to also thank my other supervisor Ras Trokovic for his supervision and support throughout this project, and for making sure that I always knew what I was doing. A big thank you also to Jere Weltner, who taught me everything about CRISPRa mediated reprogramming, and more specifically to both Ras and Jere for conducting the preliminary transgenic reprogramming experiment, so that I was able to select a cell line for the CRISPRa experiments. Another big thank you to Pauliina Paloviita, who, on top of being my go-to person with stupid questions and general life coach, helped me with the culture and derivation of all the 15 iPSC line clones. Thank you also to Heli Grym for running and analyzing my FACS samples and Anne Nyberg from THL Biobank for conducting cell line identity genotyping. I want to also thank all the other members of the lab for helping me with whatever I needed help with and making my time in the lab such a great experience. The samples used for the research were obtained from THL Biobank. We thank all study participants for their generous participation in the Psychiatric Family studies.

I am grateful for Timo, Ras, Jere and Aija Kyttälä from THL for revising my thesis and giving insightful comments on how to improve and polish it.

My friends were an inexhaustible source of peer support, without who this whole thesis experiment would have been much harder. Especially Tytti, thanks for the love and support and everything.

Finally, I want to thank my lovely family, who always support me and cheer for me, even though they have no idea what I am actually doing.

# 7. Literature


Altmann, M. et al., 2006. Transcriptional activation by EBV nuclear antigen 1 is essential for the expression of EBV's transforming genes. *Proceedings of the National Academy of Sciences*, 103(38), pp.14188–14193.

Auton, A. et al., 2015. A global reference for human genetic variation. *Nature*, 526(7571), pp.68–74.

Babcock, G.J., Hochberg, D. & Thorley-Lawson, D.A., 2000. The expression pattern of Epstein-Barr virus latent genes in vivo is dependent upon the differentiation stage of the infected B cell. *Immunity*, 13(4), pp.497–506.

Badhai, J. et al., 2009. Posttranscriptional down-regulation of small ribosomal subunit proteins correlates with reduction of 18S rRNA in RPS19 deficiency. *FEBS Letters*, 583(12), pp.2049–2053.

Banovich, N.E. et al., 2018. Impact of regulatory variation across human iPSCs and differentiated cells. *Genome Research*, 28, pp.122–131.

Barkal, A.A. et al., 2016. Cas9 functionally opens chromatin. *PLoS ONE*, 11(3), pp.1–8.

Barrett, R. et al., 2014. Reliable Generation of Induced Pluripotent Stem Cells From Human Lymphoblastoid Cell Lines. *Stem Cells Translational Medicine*, 3, pp.1429–1434.

Ben-Shushan, E. et al., 1998. Rex-1, a gene encoding a transcription factor expressed in the early embryo, is regulated via Oct-3/4 and Oct-6 binding to an octamer site and a novel protein, Rox-1, binding to an adjacent site. *Molecular and cellular biology*, 18(4), pp.1866–1878.

van den Berg, D.L.C. et al., 2010. An Oct4-Centered Protein Interaction Network in Embryonic Stem Cells. *Cell Stem Cell*, 6(4), pp.369–381.

Buganim, Y. et al., 2012. Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase. *Cell*, 150(6), pp.1209–1222.

Buganim, Y., Faddah, D.A. & Jaenisch, R., 2013. Mechanisms and models of somatic cell reprogramming. *Nature Reviews Genetics*, 14(6), pp.427–439.

Çalişkan, M. et al., 2011. The effects of EBV transformation on gene expression levels and methylation profiles. *Human Molecular Genetics*, 20(8), pp.1643–1652.

Carter, K.L. et al., 2002. Epstein-Barr Virus-Induced Changes in B-Lymphocyte Gene Expression. *Journal of Virology*, 76(20), pp.10427–10436.

Chen, A.C. & Gudas, L.J., 1996. An analysis of retinoic acid-induced gene expression and metabolism in AB1 embryonic stem cells. *Journal of Biological Chemistry*, 271(25), pp.14971–14980.

Chen, S. et al., 2015. Genome-wide CRISPR screen in a mouse model of tumor growth and metastasis. *Cell*, 160(6), pp.1246–1260.

Cheng, A.W. et al., 2013. Multiplexed activation of endogenous genes by CRISPR-on, an RNA-guided transcriptional activator system. *Cell Research*, 23(10), pp.1163–1171.

Chin, M.H. et al., 2012. Induced Pluripotent Stem Cells and Embryonic Stem Cells Are Distinguished by Gene Expression Signatures. *Cell Stem Cell*, 5(1), pp.111–123.

Choi, S.M. et al., 2011. Reprogramming of EBV-immortalized B-lymphocyte cell lines into induced pluripotent Stem cells. *Blood*, 118(7), pp.1801–1805.

Chronis, C. et al., 2017. Cooperative Binding of Transcription Factors Orchestrates Reprogramming. *Cell*, 168(3), p.442–459.e20.

Correa, C.R. & Cheung, V.G., 2004. Genetic variation in radiation-induced expression phenotypes. *American journal of human genetics*, 75(5), pp.885–90.

Doench, J.G., 2018. Am i ready for CRISPR? A user's guide to genetic screens. *Nature Reviews Genetics*, 19(2), pp.67–80. Available at: http://dx.doi.org/10.1038/nrg.2017.97.

Dominguez, A.A., Lim, W.A. & Qi, L.S., 2015. Beyond editing: repurposing CRISPR-Cas9 for precision genome regulation and interrogation. *Nature reviews. Molecular cell biology*, 17(1), pp.5–15.

Duan, S. et al., 2007. Mapping genes that contribute to daunorubicin-induced cytotoxicity. *Cancer Research*, 67(11), pp.5425–5433.

Ebrahimi, B., 2015. Reprogramming barriers and enhancers: Strategies to enhance the efficiency and kinetics of induced pluripotency. *Cell Regeneration*, 4(10), pp.1–12.

Fujimori, K. et al., 2016. Modeling neurological diseases with induced pluripotent cells reprogrammed from immortalized lymphoblastoid cell lines. *Molecular Brain*, 9(88), pp.1–14.

Garneau, J.E. et al., 2010. The CRISPR/cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature*, 468(7320), pp.67–71.

Golipour, A. et al., 2012. A late transition in somatic cell reprogramming requires regulators distinct from the pluripotency network. *Cell Stem Cell*, 11(6), pp.769–782.

Hannula, K. et al., 2001. Maternal and paternal chromosomes 7 show differential methylation of many genes in lymphoblast DNA. *Genomics*, 73, pp.1–9.

Hansson, J. et al., 2012. Highly Coordinated Proteome Dynamics during Reprogramming of Somatic Cells to Pluripotency. *Cell Reports*, 2(6), pp.1579–1592.

Holmans, P.A. et al., 2007. Genetics of Recurrent Early-Onset Major Depression ( GenRED ): Final Genome Scan Report. *The American Journal of Psychiatry*, 164(2), pp.248–258.

Hong, H. et al., 2009. Suppression of induced pluripotent stem cell generation by the p53-p21 pathway. *Nature*, 460(7259), pp.1132–1135.

Hu, K., 2014. Vectorology and Factor Delivery in Induced Pluripotent Stem Cell Reprogramming. *Stem Cells and Development*, 23(12), pp.1301–1315.

Hu, V.W. et al., 2006. Gene expression profiling of lymphoblastoid cell lines from monozygotic twins discordant in severity of autism reveals differential regulation of neurologically relevant genes. *BMC Genomics*, 7(118), pp.1–18.

Jachowicz, J.W. et al., 2017. LINE-1 activation after fertilization regulates global chromatin accessibility in the early mouse embryo. *Nature Genetics*, 49(10), pp.1502–1510.

Jinek, M. et al., 2012. A Programmable Dual-RNA – Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science*, 337(17), pp.816–822.

Jinek, M. et al., 2014. Structures of Cas9 endonucleases reveal RNA-mediated conformational

activation. *Science*, 343(6176), pp.1–11.

Jusiak, B. et al., 2016. Engineering Synthetic Gene Circuits in Living Cells with CRISPR Technology. *Trends in Biotechnology*, 34(7), pp.535–547.

Kaiser, C. et al., 1999. The Proto-Oncogene c-myc Is a Direct Target Gene of Epstein-Barr Virus Nuclear Antigen 2. *Journal of Virology*, 73(5), pp.4481–4484.

Kang, M.S., Hung, S.C. & Kieff, E., 2001. Epstein-Barr virus nuclear antigen 1 activates transcription from episomal but not integrated DNA and does not alter lymphocyte growth. *Proceedings of the National Academy of Sciences of the United States of America*, 98(26), pp.15233–15238.

Kim, J. et al., 2008. An Extended Transcriptional Network for Pluripotency of Embryonic Stem Cells. *Cell*, 132(6), pp.1049–1061.

Kleinstiver, B.P. et al., 2015. Broadening the targeting range of Staphylococcus aureus CRISPR-Cas9 by modifying PAM recognition. *Nature Biotechnology*, 33(12), pp.1293–1298.

Knaupp, A.S. et al., 2017. Transient and Permanent Reconfiguration of Chromatin and Transcription Factor Occupancy Drive Reprogramming. *Cell Stem Cell*, 21(6), pp.834–845.

Knight, S.C., Tjian, R. & Doudna, J.A., 2018. Genomes in Focus: Development and Applications of CRISPR-Cas9 Imaging Technologies. *Angewandte Chemie - International Edition*, 57(16), pp.4329–4337.

Koche, R.P. et al., 2011. Reprogramming factor expression initiates widespread targeted chromatin remodeling. *Cell Stem Cell*, 8(1), pp.96–105.

Kumar, S. et al., 2016. Utility of Lymphoblastoid Cell Lines for Induced Pluripotent Stem Cell Generation. *Stem Cells International*, 2016.

Liao, B. et al., 2011. MicroRNA cluster 302-367 enhances somatic cell reprogramming by accelerating a mesenchymal-to-epithelial transition. *Journal of Biological Chemistry*, 286(19), pp.17359–17364.

Liu, L. et al., 2014. Transcriptional pause release is a rate-limiting step for somatic cell reprogramming. *Cell Stem Cell*, 15(5), pp.574–588.

Liu, P. et al., 2018. CRISPR-Based Chromatin Remodeling of the Endogenous Oct4 or Sox2 Locus Enables Reprogramming to Pluripotency. *Cell Stem Cell*, 22(2), pp.252–261.

Liu, X. et al., 2017. Comprehensive characterization of distinct states of human naive pluripotency generated by reprogramming. *Nature Methods*, 14(11), pp.1055–1062.

Maeder, M.L. et al., 2013. CRISPR RNA–guided activation of endogenous human genes. *Nature Methods*, 10(10), pp.977–979.

Mali, P., Yang, L., et al., 2013. RNA-Guided Human Genome Engineering via Cas9. *Science*, 339(6121), pp.823–826.

Mali, P., Esvelt, K.M. & Church, G.M., 2013. Cas9 as a versatile tool for engineering biology. *Nat Methods*, 10(10), pp.957–963.

Marión, R.M. et al., 2009. A p53-mediated DNA damage response limits reprogramming to ensure iPS cell genomic integrity. *Nature*, 460(7259), pp.1149–1153.

Martí, M. et al., 2013. Characterization of pluripotent stem cells. *Nature Protocols*, 8(2), pp.223–253.

Mei, Y.P. et al., 2006. siRNA targeting LMP1-induced apoptosis in EBV-positive lymphoma cells is associated with inhibition of telomerase activity and expression. *Cancer Letters*, 232(2), pp.189–198.

Min, J.L. et al., 2010. Variability of gene expression profiles in human blood and lymphoblastoid cell lines. *BMC Genomics*, 11(96), pp.1–14.

Mitsui, K. et al., 2003. The Homeoprotein Nanog Is Required for Maintenance of Pluripotency in Mouse Epiblast and ES Cells. *Cell*, 113(5), pp.631–642.

Mojica, F.J.M. et al., 2009. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology*, 155(3), pp.733–740.

Monnier, N. et al., 2003. A homozygous splicing mutation causing a depletion of skeletal muscle RYR1 is associated with multi-minicore disease congenital myopathy with ophthalmoplegia. *Human Molecular Genetics*, 12(10), pp.1171–1178.

Nakagawa, M. et al., 2008. Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts. *Nature Biotechnology*, 26(1), pp.101–106.

NEB, 2017. New England Biolabs Tm calculator version 1.9.8. Available at: https://tmcalculator.neb.com/#!/main [Accessed May 2, 2018].

Nefzger, C.M. et al., 2017. Cell Type of Origin Dictates the Route to Pluripotency. *Cell Reports*, 21(10), pp.2649–2660.

Nichols, J. et al., 1998. Formation of pluripotent stem cells in the mammalian embryo dependes on the POU transcription factor Oct4. *Cell*, 95(3), pp.379–391.

Nie, Z. et al., 2012. c-Myc Is a Universal Amplifier of Expressed Genes in Lymphocytes and Embryonic Stem Cells. *Cell*, 151(1), pp.68–79.

Niedobitek, G., Meru, N. & Delecluse, H.J., 2001. Epstein-Barr virus infection and human malignancies. *International journal of experimental pathology*, 82, pp.149–170.

NIMH, 2018. NIMH Repository and Genomics Resource. *NIMH-RGR Data Explorer*. Available at: https://explorer.nimhgenetics.org/ [Accessed April 12, 2018].

Nishimura, K. et al., 2014. Manipulation of KLF4 expression generates iPSCs paused at successive stages of reprogramming. *Stem Cell Reports*, 3(5), pp.915–929.

Niu, N. et al., 2010. Radiation pharmacogenomics: a genome-wide association approach to identify radiation response biomarkers using human lymphoblastoid cell lines. *Genome Research*, 20, pp.1482–1492.

Okita, K. et al., 2011. A more efficient method to generate integration-free human iPS cells. *Nature Methods*, 8(5), pp.409–412.

Okita, K. & Yamanaka, S., 2011. Induced pluripotent stem cells: opportunities and challenges. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1575), pp.2198–2207.

Okubo, M. et al., 2001. Clonal chromosomal aberrations accompanied by strong telomerase activity in immortalization of human B-lymphoblastoid cell lines transformed by Epstein-Barr virus. *Cancer Genetics and Cytogenetics*, 129(1), pp.30–34.

Perez-Pinera, P. et al., 2013. RNA-guided gene activation by CRISPR-Cas9–based transcription

factors. *Nature Methods*, 10(10), pp.973–976.

Polo, J.M. et al., 2012. A molecular roadmap of reprogramming somatic cells into iPS cells. *Cell*, 151(7), pp.1617–1632.

Polo, J.M. et al., 2010. Cell type of origin influences the molecular and functional properties of mouse induced pluripotent stem cells. *Nature Biotechnology*, 28(8), pp.848–855.

Pulecio, J. et al., 2017. CRISPR/Cas9-Based Engineering of the Epigenome. *Cell Stem Cell*, 21(4), pp.431–447.

Puzzo, D. et al., 2015. Rodent models for Alzheimer's disease drug discovery. *Expert Opinion on Drug Discovery*, 10(7), pp.703–711.

Radzisheuskaya, A. & Silva, J.C.R., 2014. Do all roads lead to Oct4? The emerging concepts of induced pluripotency. *Trends in Cell Biology*, 24(5), pp.275–284.

Rahl, P.B. et al., 2010. C-Myc regulates transcriptional pause release. *Cell*, 141(3), pp.432–445.

Rajesh, D. et al., 2011. Human lymphoblastoid B-cell lines reprogrammed to EBV-free induced pluripotent stem cells. *Blood*, 118(7), pp.1797–1800.

Redon, R. et al., 2006. Global variation in copy number in the human genome. *Nature*, 444(7118), pp.444–454.

Ripke, S. et al., 2011. Genome-wide association study identifies five new schizophrenia loci. *Nature Genetics*, 43(10), pp.969–978.

Schwarz, B.A. et al., 2014. Nanog is dispensable for the generation of induced pluripotent stem cells. *Current Biology*, 24(3), pp.347–350.

Scotland, K.B. et al., 2008. Analysis of Rex1 (Zfp42) function in embryonic stem cell differentiation. *Developmental Dynamics*, 238(8), pp.1863–1877.

Shi, W. et al., 2006. Regulation of the Pluripotency Marker Rex-1 by Nanog and Sox2. *Journal of Biological Chemistry*, 281(33), pp.23319–23325.

Shi, Y. et al., 2017. Induced pluripotent stem cell technology: A decade of progress. *Nature Reviews Drug Discovery*, 16(2), pp.115–130.

Silva, J. et al., 2009. Nanog Is the Gateway to the Pluripotent Ground State. *Cell*, 138(4), pp.722–737.

Sklar, P. et al., 2011. Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. *Nature Genetics*, 43(10), pp.977–985.

Smith, Z.D. et al., 2010. Dynamic single-cell imaging of direct reprogramming reveals an early specifying event. *Nature Biotechnology*, 28(5), pp.521–526.

Son, D. et al., 2017. Generation of induced pluripotent stem cell (iPSC) line from Charcot-Marie-Tooth disease patient with MPZ mutation (CMT1B). *Stem Cell Research*, 24, pp.5–7.

Son, M.Y. et al., 2013. Unveiling the critical role of REX1 in the regulation of human stem cell pluripotency. *Stem Cells*, 31(11), pp.2374–2387.

Spender, L.C. et al., 2001. Direct and Indirect Regulation of Cytokine and Cell Cycle Proteins by EBNA-2 during Epstein-Barr Virus Infection Direct and Indirect Regulation of Cytokine and Cell Cycle Proteins by EBNA-2 during Epstein-Barr Virus Infection. *Journal of virology*, 75(8), pp.3537–3546.

Di Stefano, B. et al., 2016. C/EBPα creates elite cells for iPSC reprogramming by upregulating Klf4 and increasing the levels of Lsd1 and Brd4. *Nature Cell Biology*, 18(4), pp.371–381.

Di Stefano, B. et al., 2014. C/EBPα poises B cells for rapid reprogramming into induced pluripotent stem cells. *Nature*, 506(7487), pp.235–9.

Sugimoto, M. et al., 1999. Incorrect Use of "Immortalization" for B-Lymphoblastoid Cell Lines Transformed by Epstein-Barr Virus. *Journal of Virology*, 73(11), pp.9690–9691.

Sugimoto, M. et al., 2004. Steps involved in immortalization and tumorigenesis in human B-lymphoblastoid cell lines transformed by Epstein-Barr virus. *Cancer Research*, 64(10), pp.3361–3364.

Takahashi, J., 2018. Stem cells and regenerative medicine for neural repair. *Current Opinion in Biotechnology*, 52, pp.102–108.

Takahashi, K. et al., 2007. Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors. *Cell*, 131(5), pp.861–872.

Takahashi, K. & Yamanaka, S., 2006. Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell*, 126(4), pp.663–676.

Takahashi, T. et al., 2003. In Vitro Establishment of Tumorigenic Human B-Lymphoblastoid Cell Lines Transformed by Epstein-Barr Virus. *DNA and Cell Biology*, 22(11), pp.727–735.

Takakuwa, T. et al., 2004. Integration of Epstein-Barr Virus into Chromosome 6q15 of Burkitt Lymphoma Cell Line (Raji) Induces Loss of BACH2 Expression. *American Journal of Pathology*, 164(3), pp.967–974.

Tanenbaum, M.E. et al., 2014. A protein-tagging system for signal amplification in gene expression and fluorescence imaging. *Cell*, 159(3), pp.635–646.

Tapia, N. et al., 2015. Dissecting the role of distinct OCT4-SOX2 heterodimer configurations in pluripotency. *Scientific Reports*, 5, pp.1–9.

Theunissen, T.W. & Jaenisch, R., 2014. Molecular control of induced pluripotency. *Cell Stem Cell*, 14(6), pp.720–734.

Thomas, S.M. et al., 2015. Reprogramming LCLs to iPSCs Results in Recovery of Donor-Specific Gene Expression Signature. *PLoS Genetics*, 11(5), pp.1–17.

Töhönen, V. et al., 2015. Novel PRD-like homeodomain transcription factors and retrotransposon elements in early human development. *Nature Communications*, 6, p.8207.

Viswanathan, S.R. & Daley, G.Q., 2010. Lin28: A MicroRNA Regulator with a Macro Role. *Cell*, 140(4), pp.445–449.

Wang, Y., Bi, Y. & Gao, S., 2017. Epigenetic regulation of somatic cell reprogramming. *Current Opinion in Genetics and Development*, 46, pp.156–163.

Watters, J.W. et al., 2004. Genome-wide discovery of loci influencing chemotherapy cytotoxicity. *Proceedings of the National Academy of Sciences of the United States of America*, 101(32), pp.11809–14.

Weltner, J. et al., 2018. Human pluripotent reprogramming with CRISPR activators. *Nature Communications*.

Wen, W. et al., 2016. Enhanced generation of integration-free iPSCs from human adult peripheral

blood mononuclear cells with an optimal combination of episomal vectors. *Stem Cell Reports*, 6(6), pp.873–884.

Wu, H., Ceccarelli, D.F. & Frappier, L., 2000. The DNA segregation mechanism of Epstein-Barr virus nuclear antigen 1. *EMBO reports*, 1(2), pp.140–144.

Xiao, X. et al., 2016. Generation of Induced Pluripotent Stem Cells with Substitutes for Yamanaka ' s Four Transcription Factors. *Cellular Reprogramming*, 18(0), pp.1–17.

Yamashita, A. et al., 2014. Statin treatment rescues FGFR3 skeletal dysplasia phenotypes. *Nature*, 513(7519), pp.507–511.

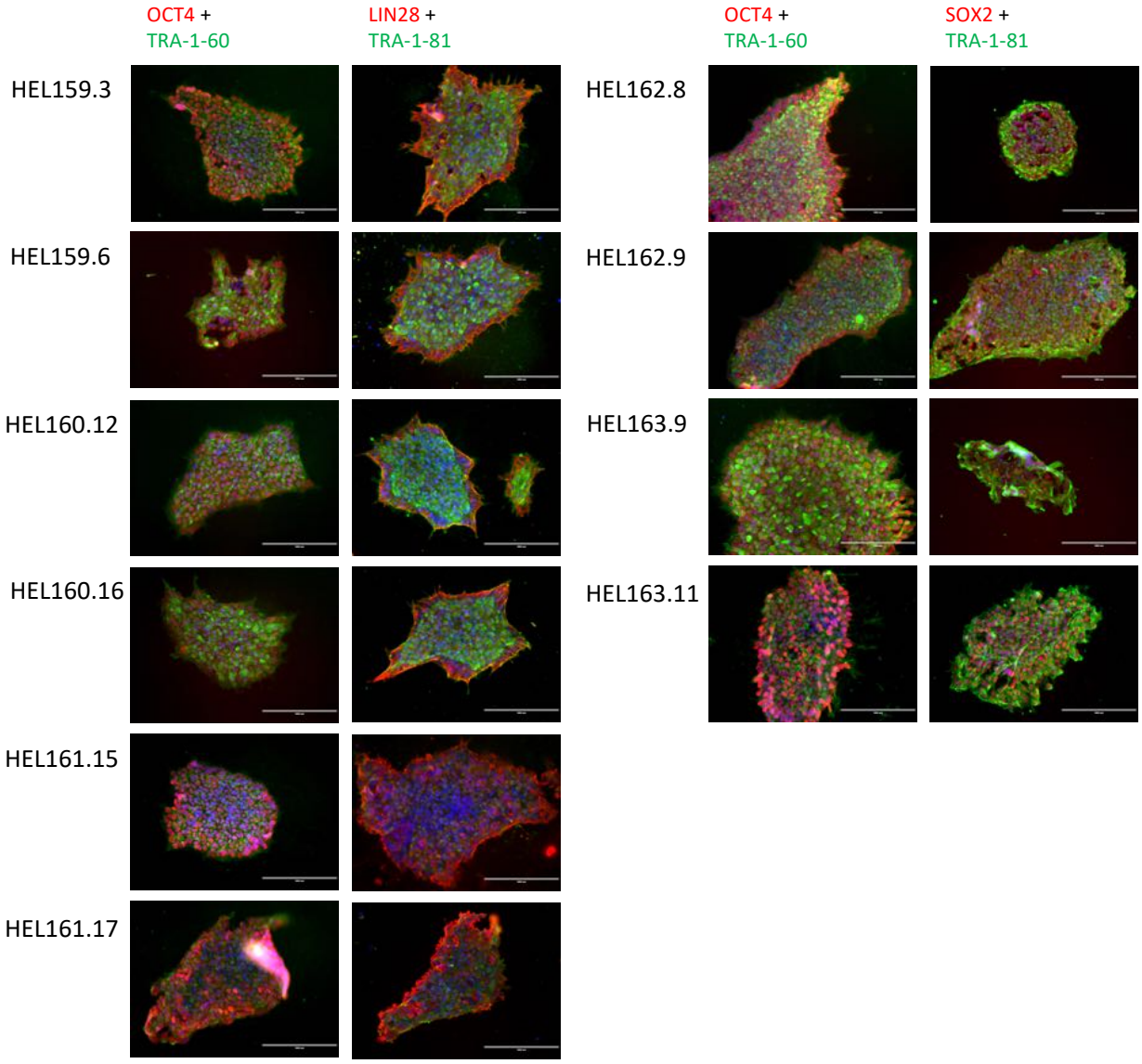
Yates, J.L., Warren, N. & Sugden, B., 1985. Stable replication of plasmids derived from Epstein-Barr virus in various mammalian cells. *Nature*, 313(6005), pp.812–815.

Zeng, Y. et al., 2016. Lin28A Binds Active Promoters and Recruits Tet1 to Regulate Gene Expression. *Molecular Cell*, 61(1), pp.153–160.

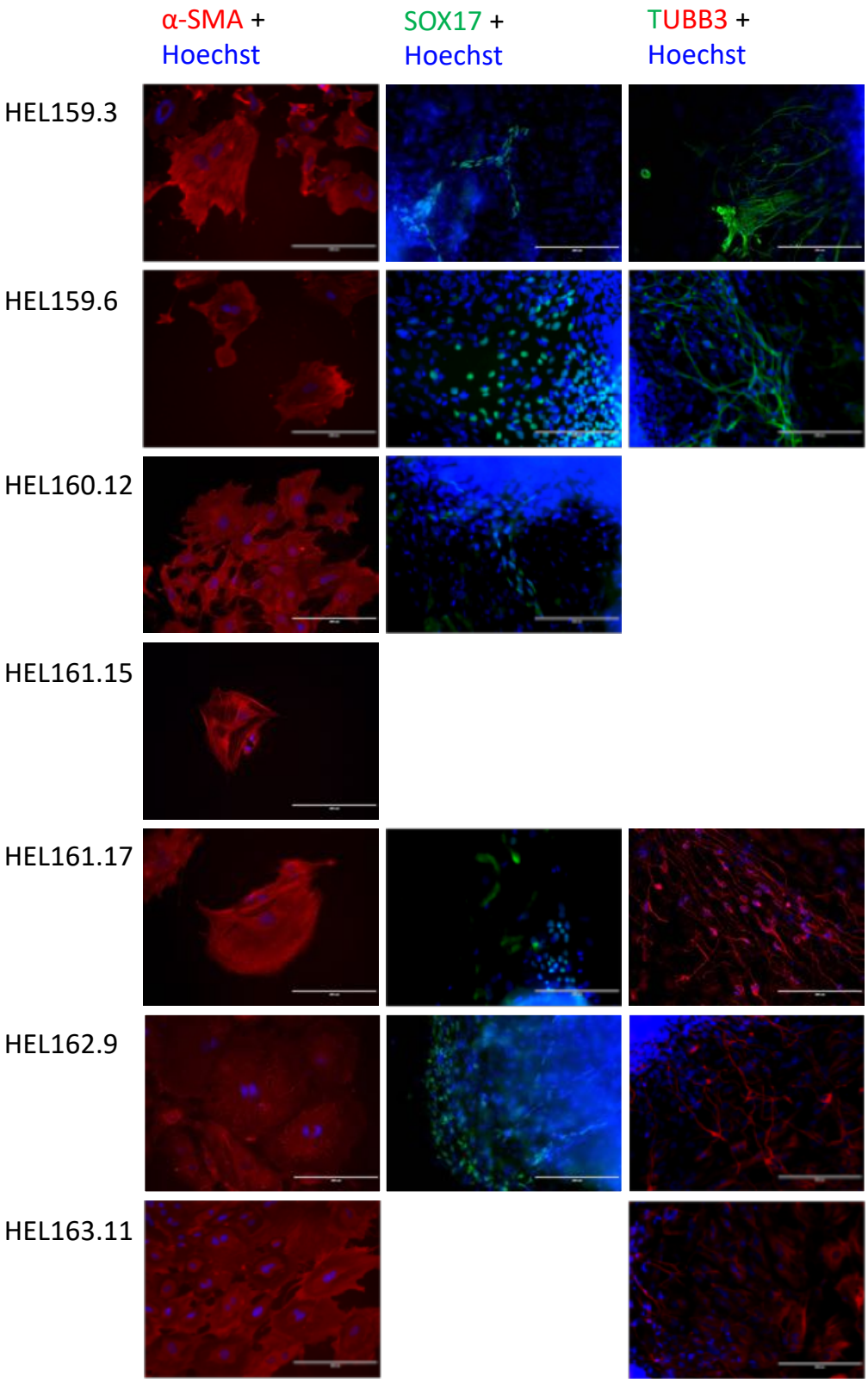
Zhang, Z. & Wu, W., 2013. Sodium Butyrate Promotes Generation of Human Induced Pluripotent Stem Cells Through Induction of the miR302/367 Cluster. *Stem Cells and Development*, 22(16), pp.2268–2277.

# 8. Appendices

## 8.1. Supplementary figures




Supplementary figure 1. Pluripotency marker expression of rest of the iPSC line clones determined by ICC. The lines were positive for the expression of OCT4, TRA-1-60, TRA-1-81, and LIN28 or SOX2. Nuclei stained with Hoechst33342. Scale bar = 200 µm.




Supplementary figure 2. ICC shows clear signal of tri-lineage differentiation of 4 iPSC lines, and differentiation to one or two germ cell layers of 3 iPSC lines. Smooth muscle actin (α-SMA) was used as an mesendodermal marker, SOX17 as a endodermal marker, and β-tubulin (TUBB3) as an ectodermal marker. Nuclei stained with Hoechst33342. Scale bar = 200 μm.